Federal Committee on Statistical Methodology Research Conference Final Program and Abstract Booklet

November 14–16, 2005

Sheraton Crystal City Hotel
1800 Jefferson Davis Highway
Arlington, VA 22202

Sponsored by:
Agency for Healthcare Research and Quality
Bureau of Economic Analysis
Bureau of Labor Statistics
Bureau of Transportation Statistics
Energy Information Administration
Environmental Protection Agency
National Agricultural Statistics Service
National Center for Education Statistics
National Center for Health Statistics
Office of Research, Evaluation and Statistics, Social Security Administration
Statistics of Income Division/Internal Revenue Service
U.S. Census Bureau

Hosted by:
Council of Professional Associations on Federal Statistics

FCSM Federal Committee on Statistical Methodology

Conference Management Committee

Nancy Kirkendall, Energy Information Administration

FCSM Research Conference Planning Committee

Bill Mockovak, Co-Chair, Bureau of Labor Statistics

Nancy Bates, Co-Chair, U.S. Census Bureau

Ana Aizcorbe, Bureau of Economic Analysis

Howard Bradsher-Fredrick, Energy Information Administration

Michael Cohen, Bureau of Transportation Statisitcs

Bill Davis, National Cancer Institute

Trena Ezzati-Rice, Agency for Healthcare Research and Quality

Dawn Haines, Social Security Administration

Charlie Hallahan, Economic Research Service

Lourdes Hartman, U.S. Census Bureau

Anna Holaus, U.S. Census Bureau

Myron Katzoff, National Center for Health Statisitcs

Steve Kaufman, Department of Education

Alesha Lewis, U.S. Census Bureau

Carol McDaniel, U.S. Census Bureau

Bob McEwen, National Agricultural Statistics Services

Samuel Posner, Centers for Disease Control and Prevention

Tammy Rib, Internal Revenue Service

Elaine Russell, U.S. Census Bureau

Brandon Shackelford, National Science Foundation

Edward Spar, Council of Professional Associations on Federal Statistics

The Federal Committee on Statistical Methodology Members

(April 2005)

Brian Harris-Kojetin, Chair, Office of Management and Budget

Wendy Alvey, U.S. Census Bureau

Lynda Carlson, National Science Foundation

Steven B. Cohen, Agency for Healthcare Research and Quality

Steven H. Cohen, Bureau of Labor Statistics

Lawrence H. Cox, National Center for Health Statistics

Robert E. Fay, U.S. Census Bureau

Ronald Fecso, National Science Foundation

Dennis Fixler, Bureau of Economic Analysis

Gerald W. Gates, U.S. Census Bureau

Barry Graubard, National Cancer Institute

Brain Greenberg, Social Security Administration

William Iwig, National Agricultural Statistics Service

Arthur Kennickell, Federal Reserve Board

Nancy Kirkendall, Energy Information Administration

Susan Schechter, Office of Management and Budget

Rolf Schmitt, Federal Highway Administration

Marilyn McMillen Seastrom, National Center for Education Statistics

Monroe Sirken, National Center for Health Statistics

Nancy Spruill, Department of Defense

Clyde Tucker, Bureau of Labor Statisitcs

Alan Tupek, U.S. Census Bureau

Katherine K. Wallman, (Champion) Office of Management and Budget

G. David Williamson, Agency for Toxic Substances and Disease Registry

Consultant

Robert Groves, Institute for Social Research, University of Michigan

Note: Papers and discussant comments will be available in early 2006 on <www.fcsm.gov>.

Federal Committee on Statistical Methodology

2005 FCSM Research Conference

The 2005 Federal Committee on Statistical Methodology (FCSM) Research Conference was initiated by the FCSM. The FCSM is an interagency committee dedicated to improving the quality of federal statistics. The committee's major goals are to:

- communicate and disseminate information on statistical practice among all federal statistical agencies,
- recommend the introduction of new methodologies in federal statistical programs to improve data quality, and
- provide a mechanism for statisticians in different federal agencies to meet and exchange ideas.

The 2005 FCSM Research Conference provides a forum for experts from around the world to discuss and exchange current research and methodological topics relevant to federal government statistical programs. Each day of the conference will offer papers on a wide range of topics including the use of advanced technologies for survey design and data collection, processing and dissemination, data mining, data warehousing and metadata, treatment of missing data, improving coverage and response rates, confidentiality and disclosure issues, record linkage, sample design and estimation, cognitive research and usability testing, and data quality.

Technical demonstrations on a variety of applications will run concurrently on the second day of the conference. Applications include new tools for conducting usability tests; an interagency-supported database of tested survey questions; Web-based documentation systems and computer-assisted interviewing software; as well as software that can be used for evaluating Website accessibility, suppressing cell values in tabular data, generating metadata, and modeling and analyzing survey data.

Sessions feature papers and demonstrations by government, private sector, and academic researchers from Brazil, Canada, France, Germany, Israel, Italy, the Netherlands, the United Kingdom and the United States of America. Paper sessions include an open discussion and some sessions include a formal discussion. Pre-conference papers will be distributed on a CD-ROM at the conference. The final papers will be posted to the FCSM Web site <www.fcsm.gov> following the conference.

In the opening plenary session, Miron Straf, the National Academies of Science, will discuss, "A National Statistical System in a Rapidly Changing World."

Final Program

Federal Committee on Statistical Methodology Research Conference

Arlington, Virginia—November 14-16, 2005

Monday (11/14)

7:30 a.m.
Registration
(Ballroom Foyer)

Coffee
(Lobby Atrium)

9:00–10 a.m.
Welcoming Remarks and PLENARY SESSION I
(Ballroom A-B)

10:00–10:30 a.m.
Break
(Ballroom Foyer)

10:30 a.m.–12 noon
CONCURRENT SESSION
II-A: II-B: II-C:
(Ballroom A) (Ballroom B) (Ballroom C)

12 noon–1:15 p.m.
Open

1:30–3:00 p.m.
CONCURRENT SESSION
III-A: III-B: III-C:
(Ballroom A) (Ballroom B) (Ballroom C)

3:00–3:30 p.m.
Break
(Lobby Atrium)

3:30–5:00 p.m.
CONCURRENT SESSION
IV-A: IV-B:
(Ballroom A) (Ballrooms B-C)

5:00–7:30 p.m.
Mixer (Cash Bar)
(Lobby Atrium)

Tuesday (11/15)

7:30 a.m.
Registration
(Ballroom Foyer)

Coffee
(Lobby Atrium)

9:00 a.m.–12:30 p.m.
Technical Demonstrations
(Crystal Rooms V and VI)

9:00–10:30 a.m.
CONCURRENT SESSION
V-A: V-B: V-C:
(Ballroom A) (Ballroom B) (Ballroom C)

10:30–11:00 a.m.
Break
(Lobby Atrium)

11:00 a.m.–12:30 p.m.
CONCURRENT SESSION
VI-A: VI-B: VI-C:
(Ballroom A) (Ballroom B) (Ballroom C)

12:30–1:45 p.m.
Open

2:00–3:30 p.m.
CONCURRENT SESSION
VII-A: VII-B: VII-C:
(Ballroom A) (Ballroom B) (Ballroom C)

Wednesday (11/16)

7:30 a.m.
Registration
(Ballroom Foyer)

Coffee
(Lobby Atrium)

9:00–10:30 a.m.
CONCURRENT SESSION
VIII-A: VIII-B:
(Ballroom A) (Ballrooms B-C)

10:30–11 a.m.
Break
(Lobby Atrium)

11:00 a.m.–12:30 p.m.
CONCURRENT SESSION
IX-A: IX-B:
(Ballroom A) (Ballrooms B-C)

Meeting Rooms:

Ballroom A-C
(2nd Floor)

Crystal Rooms V and VI
(2nd Floor)

Grand Ballroom Foyer
(2nd Floor)

Lobby Atrium
(2nd Floor)

Final Program[1]

Monday, November 14

7:30 a.m.–5:00 p.m. **Grand Ballroom Foyer**
Registration

7:30–9:00 a.m. **Lobby Atrium**
Coffee

9:00–9:10 a.m. **Grand Ballroom**
Introduction and Welcoming Remarks
Katherine Wallman (Office of Management and Budget)

9:10–10:00 a.m.
PLENARY SESSION I
A National Statistical System in a Rapidly Changing World
Miron Straf (The National Academies of Science, USA)

10:00–10:30 a.m. **Lobby Atrium**
Break

10:30 a.m.–12 noon **Ballroom A**
CONCURRENT SESSION II-A:
TIME SERIES AND STATISTICAL MODELING

Chair: Charlie Hallahan (Economic Research Service, USA)

Multivariate Statistical Modeling With Survey Data
Tihomir Asparouhov (Muthen and Muthen, USA)
Bengt Muthen (University of California, Los Angeles, USA)

Application of Concurrent Seasonal Adjustment to the CPI
Daniel Chow (Bureau of Labor Statistics, USA)
Jeff Wilson (Bureau of Labor Statistics, USA)
Adrian Thibodeau (Standard and Poor's, USA)

Model-Based Seasonal Adjustment Diagnostics
Tucker McElroy (U.S. Census Bureau)

Discussant: Stuart Scott (Bureau of Labor Statistics, USA)

Session Organizer: Charlie Hallahan (Economic Research Service, USA)

10:30 a.m.–12 noon **Ballroom B**
CONCURRENT SESSION II-B:
MEASURING DATA QUALITY

Chair: Kevin Cecco (Statistics of Income Division, Internal Revenue Service, USA)

Data Quality of the American Community Survey Across Individuals Living in Linguistically Isolated and Non-linguistically Isolated Households: A Latent Variable Model Assessment
Paul Massell (U.S. Census Bureau)–(paper presenter)
Adam Carle (U.S. Census Bureau)

SASisfying CAPI Audit Trails for Analysis
Phuc Ho (Bureau of Labor Statistics, USA)
William Chan (Bureau of Labor Statistics, USA)
Sang Kang (Bureau of Labor Statistics, USA)
Lucien Smith (Bureau of Labor Statistics, USA)
Lucilla Tan (Bureau of Labor Statistics, USA)

Assessing the Reliability of Key Measures in the National Survey on Drug Use and Health Using a Test-Retest Methodology
Joel Kennet (Substance Abuse and Mental Health Services Administration, USA)
Dicy Painter (Substance Abuse and Mental Health Services Administration, USA)
Susan Hunter (RTI International, USA)
Rebecca Granger (RTI International, USA)
Katherine Bowman (RTI International, USA)

Address-Based Sampling Versus Random Digit Dialing: Comparison of Data Quality From BRFSS Mail and Telephone Surveys
Michael Link (Centers for Disease Control and Prevention, USA)
Ali Mokdad (Centers for Disease Control and Prevention, USA)
Michael Battaglia (Abt Associates, USA)
Pamela Giambo (Abt Associates, USA)
Martin Frankel (Abt Associates, USA)

Session Organizer: Tamara Rib (Internal Revenue Service, USA)

---

[1] In the case of coauthors, the presenter is underlined.

10:30 a.m.–12 noon **Ballroom C**

CONCURRENT SESSION II-C:
BAYESIAN APPLICATIONS

Chair: Van Parsons (National Center for Health Statistics, USA)

Bayesian Networks and Complex Survey Sampling From Finite Populations
Mauro Scanu (Institute of National Statistics (ISTAT), Italy)
Marco Ballin (Institute of National Statistics (ISTAT), Italy)
Paola Vicard (University Roma Tre, Italy)

Investigation of Variance Components in the Medical Expenditure Panel Survey
Robert Baskin (Agency for Healthcare Research and Quality, USA)

A Bayesian Detection of the Onset of Activity Limitation Among Adults in NHIS
Balgobin Nandram (Worcester Polytechnic Institute, USA)
Jai Won Choi (National Center for Health Statistics, USA)

Discussant: Myron Katzoff (National Center for Health Statistics, USA)

Session Organizer: Myron Katzoff (National Center for Health Statistics, USA)

12:00 noon–1:15 p.m.
Open Lunch

1:30–3:00 p.m. **Ballroom A**

CONCURRENT SESSION III-A:
INTERNET APPLICATIONS AND SURVEY MODE COMPARISONS

Chair: Nancy Bates (U.S. Census Bureau)

Innovative Web-Based Documentation System Designed for the National Survey on Drug Use and Health
Nanthini Ganapathi (RTI International, USA)
Susan Myers (RTI International, USA)
Inga Allred (RTI International, USA)

Population Surveys From RDD Telephone and Internet Panel Samples: A Weighted Comparison of Two National Taxpayer Surveys
John Boyle (Schulman, Ronca & Bucuvalas, Inc., USA)
George Freeland (Internal Revenue Service, USA)
Lynda Mulvany (Internal Revenue Service, USA)

Potential Utility of Web-Based Data Collection Options
Heather Contrino (NuStats, USA)

Redesigning the Census Bureau's Intranet Site for Improved Usability: Methods and Lessons Learned
Elizabeth Murphy (U.S. Census Bureau)
Susan Ciochetto (U.S. Census Bureau)
Carol Bateman (U.S. Census Bureau)

Session Organizer: Nancy Bates (U.S. Census Bureau)

10:30 a.m.–12 noon **Ballroom B**

CONCURRENT SESSION III-B:
SAMPLE DESIGN AND ESTIMATION

Chair: Eric Schindler (U.S. Census Bureau)

Variance Estimation and Inference From Complex Survey Data in the Presence of Interviewer-Level Measurement Error
John Eltinge (Bureau of Labor Statistics, USA)
Moon Jung Cho (Bureau of Labor Statistics, USA)
Partha Lahiri (Joint Program in Survey Methodology, University of Maryland, USA)

Coordinating the PRN: Combining Sequential and Bernoulli-Type Sampling Schemes in Business Surveys
Ronit Nirel (Central Bureau of Statistics, Israel)
Aryeh Reiter (Central Bureau of Statistics, Israel)
Tzahi Makovky (Central Bureau of Statistics, Israel)
Moshe Kelner (Central Bureau of Statistics, Israel)

The Evolution of the Weekly Gasoline Price Survey Through Changes in Design and Frame
Paula Weir (Energy Information Administration, USA)
Pedro Saavedra (ORC Macro, USA)
Benita O'Colmain (ORC Macro, USA)

On Calibration and Non-response Adjustment for the National Compensation Survey
Alan Dorfman (Bureau of Labor Statistics, USA)
Lawrence Ernst (Bureau of Labor Statistics, USA)
Thomas Moerhle (Bureau of Labor Statistics, USA)
Steven Paben (Bureau of Labor Statistics, USA)
Chester Ponikowski (Bureau of Labor Statistics, USA)
Michael Sverchkov (Bureau of Labor Statistics, USA)

Session Organizer: Steve Kaufman (National Center for Education Statistics, USA)

10:30 a.m.–12 noon **Ballroom C**

CONCURRENT SESSION III-C:
MODEL-BASED SURVEY ESTIMATION

Chair: William Davis (National Cancer Institute, USA)

Potential Applications of Model-Assisted Estimation to Demographic Surveys in the United States
Robert Fay (U.S. Census Bureau)

New Improved Small Area Models
Swamy Paravastu (Bureau of Labor Statistics, USA)
Tamara Zimmerman (Bureau of Labor Statistics, USA)
Edwin Robison (Bureau of Labor Statistics, USA)

Smoothing Covariance Matrices From Survey Data Using Generalized Design Effects With Application to Small Area Estimation
Avinash Singh (Statistics Canada)
Yong You (Statistics Canada)
Ralph Folsom (RTI International, USA)
Akhil Vaish (RTI International, USA)

IRT and Latent Variable Modeling for Surveys With Complex Sampling Design With a Longitudinal Context: The Case of the National Longitudinal Survey of Children and Youth in Canada.
André Cyr (Statistics Canada)
Alexander Davies (Statistics Canada)

Session Organizer: William Davis (National Cancer Institute, USA)

3:00–3:30 p.m. **Lobby Atrium**
Break

3:30–5:00 p.m. **Ballroom A**

CONCURRENT SESSION IV-A:
APPLICATIONS OF SURVEY AND STATISTICAL METHODS

Chair: Rick Moser (National Cancer Institute, USA)

Data Collection From Alaskan Cruise Ships
Nelson Andrews (U.S. Environmental Protection Agency)
Don Anderson (U.S. Environmental Protection Agency)
Elizabeth Kim (U.S. Environmental Protection Agency)

Innovative Data Collection Methods in Smallpox Program Evaluation
Brian Evans (RTI International, USA)
Brian Burke (RTI International, USA)
Paul Levy (RTI International, USA)

Statistical Assessment of the Glare Issue—Human and Natural Elements
Eun-Ha Choi (National Highway and Traffic Safety Administration, USA)
Santokh Singh (National Highway and Traffic Safety Administration, USA)

Value of Physical Change in Farm Inventories: How Does One Reconcile Annual and Subannual Estimates?
Andrew Baldwin (Statistics Canada)

Session Organizer: William Davis (National Cancer Institute, USA)

3:30–5:00 p.m. **Ballrooms B–C**

CONCURRENT SESSION IV-B:
ADVANCES IN FRAME DEVELOPMENT

Chair: Bob McEwen (National Agricultural Statistics Service, USA)

Re-thinking Statistics Canada's Business Register
Eric Rancourt (Statistics Canada)
Hélène Bérard (Statistics Canada)
Stuart Pursey (Statistics Canada)

The Application of Dual System Principles to Estimate the Number of Missing Frame Elements
Howard Bradsher-Fredrick (Energy Information Administration, USA)

Modeling Which Farms Are Not Covered by a Census List Using an Area-Frame Survey
Ted Chang (University of Virginia, USA)
Phillip Kott (National Agricultural Statistical Service, USA)

Discussant: William Wigton (National Agricultural Statistics Service, USA)

Session Organizer: Bob McEwen (National Agricultural Statistics Service, USA)

5:00–7:30 p.m. **Lobby Atrium**
Mixer (Cash Bar)

Tuesday, November 15

7:30 a.m.–3:30 p.m. **Ballroom Foyer**
Registration

7:30–9:00 a.m. **Lobby Atrium**
Coffee

9:00 a.m.–12:30 p.m. **Crystal Rooms V and VI**

TECHNICAL DEMONSTRATIONS

UTE Tools to Test the Usability of Web Sites
Justin Bailey (Mind Design Systems, Inc.)
Kent Bailey (Mind Design Systems, Inc.)

InFocus and JAWS for Accessibility Testing
Lawrence Malakhoff (U.S. Census Bureau)

Q-Bank System
Kristen Miller (National Center for Health Statistics, USA)

Innovative Web-Based Documentation System Designed for the National Survey on Drug Use and Health
Inga Allred (RTI International, USA)

Blaise IS (Computer-Assisted Interviewing Software for the Web)
Jim O'Reilly (Westat, USA)

Metadata Software for NYCHANES
Debra Reed-Gillette (Centers for Disease Control and Prevention, USA)

Web-Based Factorial Survey and Statistical Analysis Software
Sam Addala (e4xchange Corporation, USA)
Matthew Hogben (Centers for Disease Control and Prevention, USA)
Vin Addala (e4xchange Corporation, USA)

ACS Automated Cell Suppression System
Gordon Sande (Sande and Associates, Canada)

Survey Data Modeling With Mplus
Tihomir Asparouhov (Muthen & Muthen, USA)

Technical Demonstration Organizer:
William Mockovak (Bureau of Labor Statistics, USA)

9:00–10:30 a.m. **Ballroom A**

CONCURRENT SESSION V-A:
VARIANCE ESTIMATION

Chair: Charlie Hallahan (Economic Research Service, USA)

A Study of the Properties of a Bootstrap Variance Estimator Under Sampling Without Replacement
Lenka Mach (Statistics Canada)
Jean Dumais (Statistics Canada)
Lauriane Robidou (University of Social Sciences, France)

An Empirical Investigation Into the Effects of Replicate Reweighting on Variance Estimates for the Annual Capital Expenditures Survey
Katherine Thompson (U.S. Census Bureau)

Investigations of BRR Variance Estimation for the Survey of Residential Alterations and Repairs (SORAR)
Laura Ozcoskun (U.S. Census Bureau)
Katherine Thompson (U.S. Census Bureau)
Quatracia Williams (ICS-RIO, Brazil)

Discussant: Phil Kott (National Agricultural Statistics Service, USA)

Session Organizer: Charlie Hallahan (Economic Research Service, USA)

9:00–10:30 a.m. **Ballroom B**

CONCURRENT SESSION V-B:
CONFIDENTIALITY

Chair: Edward Spar (Council of Professional Associations on Federal Statistics, USA)

Evolutionary Cell Selection for Interval Publication of Sensitive Cells to Protect Confidentiality
Timothy Li (Bureau of Labor Statistics, USA)
Steve Cohen (Bureau of Labor Statistics, USA)

Federal Statistical Confidentiality and Business Data: Challenges and Continuing Issues
William Seltzer (Fordham University, USA)
Margo Anderson (University of Wisconsin–Milwaukee, USA)

Privacy Principles and Data Sharing: Implications of CIPSEA for Economic Surveys Respondents
Alfred Tuttle (U.S. Census Bureau)
Diane Willimack (U.S. Census Bureau)

Discussant: Brian Harris-Kojetin (Office of Management and Budget, USA)

Session Organizer: Edward Spar (Council of Professional Associations on Federal Statistics, USA)

9:00–10:30 a.m. **Ballroom C**

CONCURRENT SESSION V-C:
GEOSPATIAL PERSPECTIVE AND ANALYSIS

Chair: William Mockovak (Bureau of Labor Statistics, USA)

The Geospatial Perception and Its Impact on the Content and Processes of a Multi-Source Data Collection
Oliva Blum (Israel Central Bureau of Statistics)

The Geospatial Distribution of Employment
Sheryl Konigsberg (Bureau of Labor Statistics, USA)
David Talan (Bureau of Labor Statistics, USA)
Richard Clayton (Bureau of Labor Statistics, USA)

Cities of Immigrants: Intraurban Mobility Patterns of Mexican Immigrants in Gateway Cities
Pamela Rogers (University of Texas at Austin, USA)

Residential and Demographic Patterns of Immigrants in Texas at a Glance: Focus on Mexican Foreign-Born Immigrants
Yann-Yann Shieh (American Institutes for Research, USA)
Pamela Rogers (University of Texas at Austin, USA)

Session Organizer: William Mockovak (Bureau of Labor Statistics, USA)

10:30–11:00 a.m. **Lobby Atrium**
Break

11:00 a.m.–12:30 p.m. **Ballroom A**

CONCURRENT SESSION VI-A:
STATISTICAL MEASURES

Chair: Nicole Nestoriak (Bureau of Economic Analysis, USA)

Total Factor Productivity Computed and Evaluated Using Multi-Step Perturbation
Baoline Chen (Bureau of Economic Analysis, USA)
Peter Zadrozny (Bureau of Labor Statistics, USA)

New Data on Business Employment Dynamics
Akbar Sadeghi (Bureau of Labor Statistics, USA)
Richard Clayton (Bureau of Labor Statistics, USA)
David Talan (Bureau of Labor Statistics, USA)

Evaluating Estimates of Labor Demand and Turnover
Charlotte Mueller (Bureau of Labor Statistics, USA)
John Wohlford (Bureau of Labor Statistics, USA)

Measurement of Reliance on Social Security Benefits
T. Lynn Fisher (Social Security Administration, USA)

Session Organizer: Ana Aizcorbe (Bureau of Economic Analysis, USA)

11:00 a.m.–12:30 p.m. **Ballroom B**

CONCURRENT SESSION VI-B:
EVALUATING THE RESPONSE PROCESS

Chair: Marc Zodet (Agency for Healthcare Research and Quality, USA)

Participation in the National Health Interview Survey: Exploring Reasons for Reluctance Using Contact History Process Data
Nancy Bates (U.S. Census Bureau)
Andrea Piani (U.S. Census Bureau)

The Effect of Interviewer Strategies on Contact and Cooperation Rates in the National Health Interview Survey
Barbara Stussman (National Center for Health Statistics, USA)
James Dahlhamer (National Center for Health Statistics, USA)
Catherine Simile (National Center for Health Statistics, USA)

Socio-Demographic Study of Telephone Survey Nonrespondents
Timothy Triplett (Urban Institute, USA)
Natalie Abi-Habib (Urban Institute, USA)

Are Two Feet in the Door Better Than One? Using Process Data to Examine Interviewer Effort and Nonresponse Bias
Kevin Wang (RTI International, USA)
Jeremy Aldworth (RTI International, USA)
Rodney Baxter (RTI International, USA)
Joe Murphy (RTI International, USA)

Session Organizer: Trena Ezzati-Rice (Agency for Healthcare Research and Quality, USA)

11:00 a.m.–12:30 p.m. **Ballroom C**

CONCURRENT SESSION VI-C:
LINKING ADMINISTRATIVE AND SURVEY DATA

Chair: Dawn Haines (Social Security Administration, USA)

The Estimation of the Italian Households' Financial Assets and Liabilities
Ivan Faiella (Bank of Italy, Italy)
Leandro D'Aurizio (Bank of Italy, Italy)
Stefano Iezzi (Bank of Italy, Italy)
Andrea Neri (Bank of Italy, Italy)

Evaluation of CPS Tax Simulation Using Administrative IRS Data
Amy O'Hara (U.S. Census Bureau)

Evaluating Respondents' Reporting of Social Security Income in SIPP Using Administrative Data
Lydia Scoon-Rogers (U.S. Census Bureau)

Discussant: Dan Kasprzyk (Mathematica Policy Research, USA)

Session Organizer: Dawn Haines (Social Security Administration, USA)

12:30–1:45 p.m.
Open Lunch

2:00–3:30 p.m. **Ballroom A**

CONCURRENT SESSION VII-A:
ADMINISTRATIVE DATA AND DATA QUALITY

Chair: Bob McEwen (National Agricultural Statistics Service, USA)

Reporting Data Quality When Survey and Administrative Data Are Combined
Julie Trépanier (Statistics Canada)
Claude Julien (Statistics Canada)
John Kovar (Statistics Canada)

Developing Error Prone Profiles Using Administrative Data With a Control Group
Pedro Saavedra (ORC Macro, USA)
Hoke Wilson (ORC Macro, USA)

Using the Canadian Address Register in the Labour Force Survey: Implementation, Results, and Lessons Learned
Claude Turmelle (Statistics Canada)
Jean-François Rodrigue (Statistics Canada)
Gavin Thompson (Statistics Canada)

Methodological Challenges in Analyzing Patient-Reported Outcomes in a Clinical Trial
Elizabeth Hahn (Evanston Northwestern Healthcare, USA)
David Cella (Evanston Northwestern Healthcare, USA)
G. Alastair Glendenning (Novartis Pharmaceuticals Corporation, USA)

Session Organizer: Bob McEwen (National Agricultural Statistics Service, USA)

2:00–3:30 p.m. **Ballroom B**

CONCURRENT SESSION VII-B:
ANALYZING AND IMPROVING RESPONSE

Chair: Steve Machlin (Agency for Healthcare Research and Quality, USA)

Earlier Versus Later Respondent Assessments on Customer Surveys: Is There a Difference?
Howard Bradsher-Fredrick (Energy Information Administration, USA)
William Weinig (Energy Information Administration, USA)

Results From Recent Experiments on Improving Response Rates
Ron Fecso (National Science Foundation, USA)

Efficient Estimation of Response Rates When a Small Subsample of Nonrespondents Is Selected for Follow-Up Conversion
Avinash Singh (Statistics Canada)
Vincent Iannacchione (Research Triangle Institute International, USA)
Shijie Chen (RTI International, USA)
Jill Dever (Joint Program in Survey Methodology, University of Maryland, USA)

Tests of Methods to Improve Response to Physician Surveys
David Woodwell (National Center for Health Statistics, USA)
Catharine Burt (National Center for Health Statistics, USA)

Session Organizer: Trena Ezzati-Rice (Agency for Healthcare Research and Quality, USA)

2:00–3:30 p.m. **Ballroom C**

CONCURRENT SESSION VII-C:
SURVEY INTEGRATION AND PLANNING

Chair: Polly Phipps (Bureau of Labor Statistics, USA)

Integrating the Canadian Annual Survey of Manufactures Into the Unified Enterprise Survey: Challenges and Issues
Isabelle Marchand (Statistics Canada)
Emmanuel Benhin (Statistics Canada)
Jose Gaudet (Statistics Canada)

Implementing and Improving Quality Standards at the U.S. Census Bureau
John Bushery (U.S. Census Bureau)
Pamela McGovern (U.S. Census Bureau)

Developing a Generic Survey
Marla Smith (U.S. Environmental Protection Agency)
Jan Matuszko (U.S. Environmental Protection Agency)

Integrated Redevelopment of the ONS's Business Surveys and Business Register Through the Business Surveys Integration Project and Register Re-engineering Project
Mark Pont (Office for National Statistics, United Kingdom)
Pam Tate (Office for National Statistics, United Kingdom)
Paul Smith (Office for National Statistics, United Kingdom)
John Perry (Office for National Statistics, United Kingdom)

Session Organizer: William Mockovak (Bureau of Labor Statistics, USA)

Wednesday, November 16

7:30–11:00 a.m. **Ballroom Foyer**
Registration

7:30–9:00 a.m. **Lobby Atrium**
Coffee

9:00–10:30 a.m. **Ballroom A**

CONCURRENT SESSION VIII-A:
QUALITATIVE EVALUATION

Chair: Alethea Jennings (Energy Information Administration, USA)

Determining Best Practices for Usability Test Methodology: A Comparison of Assessment Methods
Elizabeth Dean (RTI International, USA)
Michael Schwerin (RTI International, USA)
Kimberly Aspinwall (RTI International, USA)

Cross-Cultural Issues in Survey Translation: Translation of Meaning and Meaning of Translation
Daniel Geller (ORC Macro, USA)
Andrey Vinokurov (ORC Macro, USA)
Tamara Martin (U.S. Department of State)

Using Affective Imagery to Understand the Quality of Survey Response
Patricia Gwartney (University of Oregon, USA)

Does Qualitative Evaluation Measure Up?
Kathleen O'Connor (National Center for Health Statistics, USA)

Session Organizer: Howard Bradsher-Fredrick (Energy Information Administration, USA)

9:00–10:30 a.m. **Ballrooms B–C**

CONCURRENT SESSION VIII-B:
PRETESTING AND USE OF NON-ENGLISH LANGUAGE QUESTIONNAIRES

Chair: Florina Serbanescu (Centers for Disease Control and Prevention, USA)

Adapting Cognitive Interview Techniques for Use in Pretesting Spanish Language Survey Instruments
Patricia Goerman (U.S. Census Bureau)

Development of Guidelines on the Use of Interpreters in Survey Interviews
Yuling Pan (U.S. Census Bureau)

The Use of Cognitive Interviewing and Behavior Coding to Evaluate Non-English Language Survey Questions: Lessons Learned
Gordon Willis (National Cancer Institute, USA)

Old Friends and Old Problems: Implementing Frequently-Used Question Types in the Cross-National Context
Janet Harkness (ZUMA, Germany)

Session Organizer: Sam Posner (Centers for Disease Control and Prevention, USA)

10:30–11:00 a.m. **Lobby Atrium**
Break

11:00 a.m.–12:30 p.m. **Ballroom A**

CONCURRENT SESSION IX-A:
USE OF INNOVATIVE TECHNOLOGIES

Chair: Shawna Waugh (Energy Information Administration, USA)

Web-Based Factorial Survey and Statistical Analysis Software
Sam Addala (e4xchange Corporation, USA)
Vin Addala (e4xchange Corporation, USA)
Matthew Hogben (Centers for Disease Control and Prevention, USA)

Protecting Sensitive Tabular Data by Complementary Cell Suppression–Myth & Reality
Ramesh Dandekar (U.S. Department of Energy)

Digital Capture of Geographic Feature Data for Surveys
Sarah Nusser (Iowa State University, USA)

Guaranteed Controlled Rounding for Many Totals in Multi-Way and Hierarchical Tables
Gordon Sande (Sande & Associates, Canada)

Session Organizer: Howard Bradsher-Fredrick (Energy Information Administration, USA)

11:00 a.m.–12:30 p.m. **Ballrooms B–C**

CONCURRENT SESSION IX-B:
IMPROVING RESPONSE AND DATA QUALITY IN ESTABLISHMENT SURVEYS

Chair: Dennis Fixler (Bureau of Economic Analysis, USA)

Applying Knowledge of Business Surveys Response Processes to the Design of Data Collection Software at the U.S. Census Bureau
Amy Anderson (U.S. Census Bureau)
Rebecca Morrison (U.S. Census Bureau)

The Response Process Model in Business Surveys: Lessons Learned by Using a Multi-Method Approach
Deirdre Giesen (Statistics Netherlands)
Tony Hak (Erasmus University, The Netherlands)

Improving Industry Descriptions for the Annual Refiling Survey
Karen Goldenberg (Bureau of Labor Statistics, USA)
Monica Dashen (Bureau of Labor Statistics, USA)

Towards Reducing Error in an Establishment Survey Through Instrument Design: Identifying the Desired Navigational Path
Cleo Redline (National Science Foundation, USA)

Session Organizer: Ana Aizcorbe (Bueau of Economic Analysis, USA)

Abstract Booklet

This section represents abstracts received as of September 1, 2005.

The following abstracts have not been edited for content.

CONCURRENT SESSION II-A:
TIME SERIES AND STATISTICAL MODELING

Multivariate Statistical Modeling With Survey Data
Tihomir Asparouhov (Muthen and Muthen, USA) and Bengt Muthen (University of California, Los Angeles, USA)

We describe an extension of the pseudo maximum likelihood (PML) estimation method developed by Skinner (1989) to multistage stratified cluster sampling designs, including finite population and unequal probability sampling. We conduct simulation studies to evaluate the performance of the proposed estimator. The estimator is also compared to the general estimating equation (GEE) method for linear regression implemented in SUDAAN. We investigate the distribution of the likelihood ratio test (LRT) statistic based on the pseudo log-likelihood value and describe an adjustment that gives correct chi-square distribution. The performance of the adjusted LRT is evaluated with a simulation study based on the Behrens-Fisher problem in a stratified cluster sampling design.

Application of Concurrent Seasonal Adjustment to the CPI
Daniel Chow, and Jeff Wilson (Bureau of Labor Statistics, USA) and Adrian Thibodeau (Standard and Poor's, USA)

This paper discusses the findings, observations, and conclusions about revision gains resulting from concurrent adjustment. The CPI program of the Bureau of Labor Statistics currently revises and produces seasonal data on an annual basis. Due to advances in computing power and software making once prohibitive and resource-intensive methodologies more feasible, concurrent seasonal adjustment is being adopted in more countries than in the past. A key reason for the growing interest in concurrent adjustment is that it results in faster and more accurate revisions to final seasonally adjusted data compared to forward factor adjustment. This produces a better indication of the trend of a time series on a more frequent basis compared to the forward factors method which generates revisions only once per year. Based on research conducted by the Australian Bureau of Statistics (ABS) which found positive revision gains for quarterly data compared to forward factors, this paper studies whether BLS' monthly CPI data exhibit similar gains. Several CPI series with varying degrees of seasonality are examined. The concurrent and forward factor adjustments are then simulated and compared using the ABS' X-11 based seasonal adjustment software, SEASABS. The findings indicate that most of the simulated CPI data do experience general revision improvements and that individual results vary depending on the volatility of the given series.

Model-Based Seasonal Adjustment Diagnostics
Tucker McElroy (U.S. Census Bureau)

Findley, McElroy, and Wills (2004) explored a diagnostic for the quality of seasonal adjustments in a model-based setting. This diagnostic measures the amount of variation in the estimated irregular component, and is a modification of SEATS' existing diagnostic, which was introduced in Maravall (2003). Findley et al. (2004) tested their diagnostic in both Monte Carlo and empirical studies by fitting data to specific SARIMA models, such as the airline model. This paper discusses the issues surrounding the implementation of the diagnostic in the Ox programming language, utilizing the SsfPack function suite. Here we treat six diagnostics, based on estimates of the trend, seasonal, irregular, and combinations thereof.

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

CONCURRENT SESSION II-B:
MEASURING DATA QUALITY

Data Quality of the American Community Survey Across Individuals Living in Linguistically Isolated and Non-linguistically Isolated Households: A Latent Variable Model Assessment
Adam Carle (U.S. Census Bureau)

To address internal validity, one aspect of data quality among many, confirmatory factor analyses for ordered-categorical measures (CFA-OCM) explored measurement invariance across linguistic isolation (LI) status in the American Community Survey. Analyses examined whether the full set of measurement parameters for the six items measuring disability demonstrated differential item functioning. Given the set of adopted statistical criteria, results generally demonstrated measurement equivalence. Findings support uniformity in the internal validity of ACS data collected across LI individuals for these items, and dispute concerns that LI substantially affects the ability of LI individuals to answer these items on the ACS in a meaningful way as compared to non-LI

individuals. When making comparisons regarding disability, investigators using ACS data can be less concerned that LI status impacts the validity of comparisons and can place greater faith in the quality of their comparisons.

SASifying CAPI Audit Trails for Analysis

Phuc Ho, William Chan, Sang Kang, Lucien Smith, and Lucilla Tan (Bureau of Labor Statistics, USA)

With the advance in portable computing resources in recent years, many data collection agencies have moved from a paper-and-pencil instrument to Computer Assisted Personal Interview (CAPI). With CAPI came the ability to capture timing and the sequence of keystrokes associated with an interview, via an audit trail. The audit trails contain a wealth of information for instrument diagnostics and survey methodologists. However, the analysis of audit trails has been limited because of the overhead in processing these detailed, semi-structured text files.

This paper reports on the development of a prototype to analyze the CAPI audit trails for the Consumer Expenditure Survey at the Bureau of Labor Statistics. The transformation of the audit trails into SAS data sets expands the accessibility to the information contained in the audit trails to a wider range of users, as programming queries on the SAS data set is simpler than programming queries on the audit trails in their native format. Techniques using SAS to parse the audit trail files into a more structured format, and examples of information gleaned from analyzing the audit trails are presented here.

Assessing the Reliability of Key Measures in the National Survey on Drug Use and Health Using a Test-Retest Methodology

Joel Kennet and Dicy Painter (Substance Abuse and Mental Health Services Administration, USA), Katherine Bowman, Rebecca Granger, and Susan Hunter (RTI International, USA)

The National Survey on Drug Use and Health (NSDUH) is a major source of information on substance use and mental illness prevalence in the United States. It is administered in households to approximately 67,500 individuals annually using a complex, multi-stage sampling design. Assessing the reliability of estimates produced by the NSDUH is of primary importance to those who use these data for research and in the making of policy decisions. In Quarters 1 and 2 of 2005, a pretest was carried out, in which approximately 200 NSDUH respondents were re-interviewed in an effort to fine tune the methods to be used in conducting a large-scale reliability field test in 2006. This paper will discuss the design and procedural considerations that were taken into account in planning the pretest and upcoming field test. These considerations included time interval between test and re-test, sample size needed for reliability estimates of low-prevalence behaviors, whether sample would be embedded vs. not embedded in the NSDUH main study, using the same vs. different interviewers for the re-interview, increased risk of loss of respondent privacy due to the provision of re-contact information, amount of incentive for the re-interview, and others. In addition, preliminary findings from the pretest that may influence methods employed in the 2006 field test, such as response rates on the re-interview, and respondent feedback, will be presented and discussed.

Address-Based Sampling Versus Random Digit Dialing: Comparison of Data Quality from BRFSS Mail and Telephone Surveys

Michael Link and Ali Mokdad (Centers for Disease Control and Prevention, USA), Michael Battaglia, Pamela Giambo, and Martin Frankel (Abt Associates, USA)

As part of an on-going effort to expand coverage for and participation in the Behavioral Risk Factor Surveillance System (BRFSS), a pilot study was conducted to determine if a mail survey conducted with a random sample of adults selected from an address-only sampling frame could rival the quality of the data collected using more traditional random digit dialed (RDD) methods. As one of the largest, RDD-based health surveys, the BRFSS is an ongoing surveillance designed to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and

preventable infectious diseases in the adult population. Moving from a telephone-only approach to one which utilizes alternative survey modes and sampling frames has the potential to improve response rates and increase the validity and reliability of the estimates obtained. This paper reports the results of a pilot study conducted in six states. Estimates derived from mail and telephone versions of the BRFSS are compared across 10 key health conditions and risk behaviors, including diabetes, high blood pressure, obesity, alcohol use, and HIV testing. Other assessments of data quality are also presented, including frame coverage, unit and item nonresponse, and alternative methods of within household randomization.

CONCURRENT SESSION II-C:
BAYESIAN APPLICATIONS

Bayesian Networks and Complex Survey Sampling From Finite Populations
Mauro Scanu and Marco Ballin (Institute of National Statistics (ISTAT), Italy) and Paola Vicard (University Roma Tre, Italy)

We propose a novel methodology based on Bayesian Networks (see Cowell et al, 1999) for the estimation of the joint probability distribution of a set of variables when samples are drawn according to complex survey designs. Bayesian Networks are widely used graphical models for the analysis in many scientific context, especially Artificial Intelligence and multivariate statistics. Variables related to the different units in the sample are currently assumed to be i.i.d. when estimating and using Bayesian Networks, and they have been never defined in the context of finite population sampling. We show that the Bayesian Networks machinery can be easily adapted to complex sampling designs through the definition of an additional node, S, representing the sample design. S is a categorical variable assuming as many categories as the different first order inclusion probabilities. The marginal probability attached to S is defined through the sampling design as the fraction of the overall weight relative to the units of the population in each S category. It will be shown that the Bayesian Network inferential engine is consistent with the usual estimators of frequencies in a finite population context. Furthermore, it will be seen how this representation suggests a number of solutions for different statistical problems. These are based on one of the most important characteristics of Bayesian Networks: an efficient propagation of information through the network. This aspect allows the formulation of solutions for the following problems: evaluation of possible scenarios helpful for policy makers; use of auxiliary information, whenever available, for the definition of more efficient estimators, as alternative calibration estimators; possibility to update/integrate coherently sets of different sample surveys. The definition of Bayesian Networks in a finite population context will also allow the use of appropriate imputation methodologies, as the one in Di Zio et al. (2004).

Investigation of Variance Components in the Medical Expenditure Panel Survey
Robert Baskin (Agency for Healthcare Research and Quality, USA)

The Medical Expenditure Panel Survey (MEPS), a national probability sample survey sponsored by the Agency for Healthcare Research and Quality, is designed to provide annual national and regional estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. MEPS has a complex sample design which makes estimation of variance components somewhat problematic. The purpose of this study is to begin an assessment of methods of estimating the variance components associated with selected variables in MEPS. Because of the scope of the work, initially the evaluation will be conducted for only two specific health related variables: healthcare expenditures and SF-12 health status measure.

A Bayesian Detection of the Onset of Activity Limitation among Adults in NHIS
Balgobin Nandram (Worcester Polytechnic Institute, USA) and Jai Won Choi (National Center for Health Statistics, USA)

Data from the 1996 National Health Interview Survey (NHIS) indicate that, due to chronic conditions, the onset of activity limitation typically occurs between age 35-55 years (i.e., the proportion of young adults with activity limitation is small and roughly constant with age and then it starts to change, roughly increasing). We use a Bayesian hierarchical model to detect the change point of a positive activity limitation status (ALS), and we include a similarity among states (small areas) in the NHIS. This is the beta-binomial model which incorporates a constant proportion until the change point with variable proportions thereafter. We use the Gibbs sampler to fit the model, and a computation of the marginal likelihoods, using an output analysis from the Gibbs sampler, provides the posterior distribution of the change point. The results show that the overall onset of

ALS is about 38 years. However, the onset is later among white adults than non-white, and substantially later among males than females. In addition, white males have the latest onset of ALS, which is significantly different from white females, and non-white males and females.

CONCURRENT SESSION III-A:
INTERNET APPLICATIONS AND SURVEY MODE COMPARISONS

Innovative Web-Based Documentation System Designed for the National Survey on Drug Use and Health

Nanthini Ganapathi, Susan Myers, and Inga Allred (RTI International, USA)

Data documentation is equally important to the success of a study as the data themselves. It is imperative that data collected and created within studies be thoroughly documented so that they can be analyzed. Developing high quality metadata is a tough challenge for any information management community. When this metadata spans several study years and involves a large volume of data, the need arises for a centralized location allowing users to add, modify, and track variables.

The National Survey on Drug Use and Health (NSDUH) is a nationwide survey conducted annually by RTI International for the Substance Abuse and Mental Health Services Administration (SAMHSA). It collects interview data from approximately 67,500 respondents in all 50 states and the District of Columbia. The NSDUH is the primary source of information on the prevalence, patterns, and consequences of alcohol, tobacco, and illegal drug use and abuse in the general U.S. civilian non-institutionalized population, age 12 and older. There are currently multiple data analyses active at different stages on the NSDUH, and each one may have as many as 3,000 variables to document.

RTI International has developed Web based Variable Tracking System (VTS) software for variable documentation and reporting for the NSDUH project. The goal was to create software that would keep current documentation available to analysts at all times and automate the creation of electronic and paper codebook. VTS is an efficient and user friendly system allowing users to enter, edit and search for information about variables. VTS also allows users to identify the individuals responsible for the variable and its documentation, so that the responsible person can be contacted for questions. In addition, this system serves as the source of all metadata that are assimilated into deliverable documentation for the client. This paper focuses on the key features of VTS and discusses its future direction.

Population Surveys From RDD Telephone and Internet Panel Samples: A Weighted Comparison of Two National Taxpayer Surveys

John Boyle (Schulman, Ronca & Bucuvalas, Inc., USA) and George Freeland and Lynda Mulvany (Internal Revenue Service, USA)

A nationwide telephone survey of 2,000 randomly-selected households was conducted in January–February 2004 to determine taxpayer use and perceptions of the paper and electronic versions of Publication 17 and the IRS Website from which the electronic version can be obtained. The survey was conducted for the Tax Forms and Publications (TFP) Division of the Internal Revenue Service (IRS) by Schulman, Ronca, and Bucuvalas, Inc. (SRBI). This telephone survey provides estimates of taxpayer characteristics, experiences, attitudes and preferences based upon a national probability sample.

As part of the qualitative research phase of this project, a Web-based survey was also conducted among a national Internet panel. A total of 9,000 invitations were e-mailed to a national sample of adult panelists. Nearly 1,000 qualified responses were received from taxpayers to the Internet survey. The response rate was reported as standard for a Web-based survey of this length by the organization who managed the panel.

Since many of the questions were identical for the telephone and Internet samples of taxpayers, a comparison of the findings between the two surveys should illuminate the implications of these two modes of data collection for surveys of taxpayers. Since approximately ninety percent of adults in the United States meet the study definition of taxpayers, these findings should also be useful to other national population surveys.

For this paper, both samples have been restricted to the person in the household most familiar with 2002 federal income tax, and the Internet sample was weighted to the parameters of the telephone sample on age, gender and education to correct for known population biases in Internet surveys. Nonetheless, many significant differences persist in other characteristics, attitudes and behaviors of the weighted Internet sample and the national telephone sample. Some of these differences may enhance the value of Internet panels in qualitative research, but reinforce existing concerns about its utility for population estimates.

Potential Utility of Web-Based Data Collection Options

Heather Contrino (NuStats, USA)

Current survey programs across the United States are struggling with issues of frame coverage and nonresponse. In an environment of limited resources, the research community is putting a great deal of effort into the investigation of frame, contact, and data collection options. Each traditional data collection mode, regardless of the sample frame, has benefits and limitations in terms of non response levels, measurement error, and other forms of bias. In recent years, there has been an increased propensity to utilize multi-mode designs as a more respondent friendly approach to increasing overall and subgroup response rates. One data collection mode that has become increasingly employed in the private sector as a supplemental data collection option is the Internet.

Utilizing data collected in the USPS Household Diary Study (HDS), this paper provides an initial examination of the potential effectiveness of the incorporation of Web based data collection options in surveys. The USPS HDS is a national study that has been conducted since 1987 and utilizes an address frame with both telephone and mail based data collection modes. The first stage of the study collects information on mail behavior and household demographics. The second stage requires the household to record all mail sent and received. Data from the second stage is retrieved via mail back diaries. In 2003, a Web based data collection option was incorporated into the research design. The survey design of the HDS includes an advance mailing to all sampled households which requests participation in the study and provides the household with the option of completing the first stage of the study via a secure study Website. Currently 10 percent of participating households complete the first phase of this two stage study via the Web (n=950).

The research design of the HDS and level of Web response provides a unique research opportunity. This paper will first present a comparative analysis of the demographic characteristics of first and second stage responders across the phone, mail, and Web data collection modes. Using household information and completion data from both the first and second phases, inferences on potential sources and impact of nonresponse will also be examined. In addition, comparisons of key estimates on mail usage and behavior will be conducted across modes. These comparative analyses will support increased information for three key research questions:

1. What are the demographic characteristics of households most likely to choose Web, phone, and mail data collection modes when provided with participation options in stage one of the study?
2. Is the level of nonresponse in the second stage of the study (mail back only) different across stage one completion mode groups (Web, phone, mail)?
3. Are there significant differences in mail behavior across stage one completion mode groups? How do these differences, if any, impact overall key estimates of mail volumes?

Redesigning the Census Bureau's Intranet Site for Improved Usability: Methods and Lessons Learned

Elizabeth Murphy, Susan Ciochetto, and Carol Bateman (U.S. Census Bureau)

Redesign of the U. S. Census Bureau's intranet site was undertaken to increase the ease of finding information on the site. As previously designed, the site reflected the organizational structure of the Census Bureau, but new employees did not necessarily know which organization was responsible for what information. The objective was to redesign the site using functional, logical categories for the information content. This presentation will focus on the design and evaluation methods that were applied to the redesign effort, including card sorting and eye tracking. These methods can be used in any Web-site design or re-design context.

CONCURRENT SESSION III-B:

SAMPLE DESIGN AND ESTIMATION

Variance Estimation and Inference From Complex Survey Data in the Presence of Interviewer-Level Measurement Error

John Eltinge and Moon Jung Cho (Bureau of Labor Statistics, USA) and Partha Lahiri (Joint Program in Survey Methodology, University of Maryland, USA)

In the analysis of data collected through a complex sample design, one generally needs to account for variability associated both with customary sampling errors and with measurement errors. For cases in which measurement errors are attributable to individual sample elements (e.g., sample households or sample establishments), it is relatively simple to extend standard design-based methods for variance estimation and inference. However, such extensions become more complex for cases in which a substantial portion of

measurement error effects are attributable to interviewers, rather than to sample elements. Following a review of previous literature in the area, this paper focuses on three inter-related problems. First, we consider stratified multistage sample designs in which some interviewers collect data in more than one primary sample unit (PSU). We develop a simple collapse-based variance estimator which is conservative under mild regularity conditions. In addition, we discuss related inference issues involving confidence sets and coverage rates. Second, we extend the first case to cases in which measurement error variances may be associated with specified characteristics of the sample units (e.g., household or establishment size or composition) and of the interviewers (e.g., experience or workload). Third, we use a hierarchical model to evaluate some properties of the first two sets of methods, and to develop some alternative variance estimators. The principal ideas are motivated by, and illustrated with, a specific application to the Consumer Expenditure Interview Survey.

Coordinating the PRN: Combining Sequential and Bernoulli-Type Sampling Schemes in Business Surveys

Ronit Nirel, Aryeh Reiter, Tzahi Makovky, and Moshe Kelner (Central Bureau of Statistics, Israel)

Business populations are constantly changing because of births, deaths, splits, merges and changes in size and activity. To estimate efficiently changes over time, panel designs that maximize overlap between samples are frequently used. In addition, samples should be updated periodically to reflect the dynamics of the population. Permanent Random Numbers (PRNs) are a useful tool for coordinating samples over time while preserving considerable overlap between successive samples. Each unit in the frame is assigned a random number in the interval [0,1] that is permanently associated with that unit. In preparation for the sampling, the units are sorted by their PRN. Here we deal with two sampling schemes that use PRNs to select n units from a population of size N. In a "sequential" scheme the sample comprises the first n units in the sorted list. The upper bound of the PRNs of the sampled units is thus random. In a Bernoulli scheme, the sample is obtained by selecting all units with a PRN not larger than n/N. The sample size in the Bernoulli scheme is therefore random.

In Israel, we are redesigning the samples redesign for the business surveys. Several principles guide the new design. First, businesses are stratified by industry and size, and units are selected in each stratum with equal probabilities. Second, at the beginning of each year an annual base-sample is reallocated. The resulting sample should have a sizable overlap with the sample of the previous year. Third, the sample is updated once in two months to account for changes in the frame. Last, it is required that for each stratum the sampling weight is kept constant at its base-sample value throughout the year. Using a sequential PRN scheme for the base-sample complies with the first two principles. As for the bi-monthly updates, a Bernoulli scheme satisfies the requirement of constant weights over time. However, using an unadjusted Bernoulli scheme creates two problems: (a) the cumulative sample size over updates may have large deviations from its expected value; and (b) because of the differences between the sequential and Bernoulli schemes, the upper bound of PRNs of units that are added during the year may differ from the base-sample value. Hence, if no corrective action is taken, their selection probabilities in the next base-sample will be different from those of the old units.

In this paper we suggest a sampling method that controls the overall sample size of the five inter-annual updates, and keeps the expected weights unchanged over time. The method combines collocated sampling with randomization of conditional inclusion probabilities. We also propose adjusted PRNs for units included in the inter-annual frames that have two desired properties: they have a uniform distribution in the unit interval, and the maximal value for the sampled units is the same as for the base-sample. These adjusted PRNs are used for the next reallocation. We show that an intuitive adjustment based on a simple rescaling of the $[0,n/N]$ interval to an $[0, \pi]$ interval (where π is the maximal PRN for the base-sample units) results in PRNs that have a distribution which is a mixture of Beta distributions rather than a uniform distribution. Using this result we propose a correction for the rescaled PRNs that has the desired properties. Finally, we illustrate the method for simulated data and for data from the Manufacturing Indices Survey.

The Evolution of the Weekly Gasoline Price Survey Through Changes in Design and Frame

Paula Weir (Energy Information Administration, USA) , Pedro Saavedra, and Benita O'Colmain (ORC Macro, USA)

The Energy Information Administration (EIA) weekly survey of gasoline prices produces estimates of gasoline pump prices at the national and regional levels, as well as estimates for several states and cities, two formulations and three grades of gasoline. This survey originated as a response to the First Gulf War and was later expanded to monitor the effects of the Clean Air Act on the price of gasoline. The original design was a two-phase sample, with a monthly survey's sample of gasoline refiners and resellers serving as the first phase, and the individual gasoline stations owned by the monthly respondents serving as the second. Subsequent expansions to produce estimates at lower geographic levels required the use of two cycles of the monthly survey sample and the U.S. Census Bureau County Business Patterns database. The most recent expansion of state and city estimates, however, required a new approach to the sampling frame and sample design. The new design made use of an outlet level frame, allocated stations to counties, and then sampled stations from the

selected counties. However, because the old corporate level frame and the new outlet level frame approached the population differently, the potential existed for discontinuity in the survey estimates. In order to attenuate the discontinuity, the sample drawn under the new design was combined with a subsample of the old design. Variance estimates of the new sample were obtained taking into account the procedure that combined the two samples and frames.

On Calibration and Non-response Adjustment for the National Compensation Survey
Alan Dorfman, Lawrence Ernst, Thomas Moerhle, Steven Paben, Chester Ponikowski, and Michael Sverchkov
(Bureau of Labor Statistics, USA)

Most surveys have to deal with non-response. One of the approaches that is used to compensate for this is the correction of the initial weights of respondents through an adjustment factor. Under this approach, the entire sample is divided into cells of similar units, for example, representatives of a particular area, establishment and occupation. For each usable unit (i.e. a unit that contains all necessary information for the future analysis) its frame sample weight is multiplied by the ratio of weighted estimates of the viable units in the cell to weighted estimates of all the usable units in the same cell (non-response adjustment factor). In other words, assuming that the non-respondents in every cell are of the same type as the respondents, the weights of the first are transferred to the last such that the survey estimates of respondent totals coincides with the survey estimates of population totals.

Another simple approach for dealing with non-response is to use post-stratification (benchmarking to known population subtotals). If population totals are known for some subsets of a population (population cells) then instead of correcting the sampling weights for non-response adjustment factors one can correct the sample weight of a usable unit by a ratio of the known total in the cell to its weighted estimate. The advantage of this kind of adjustment is that the weighted benchmark estimator estimates population totals for cells used in the estimation process exactly and is expected to give more accurate results for other kinds of estimates based on weights. Another advantage of this method is that it allows the use of "fresh" information. Very often surveys use the same sampling frame for years even though the frame changes over time, which can lead to biased estimates. Benchmarking makes a correction of the older weights without additional sampling.

Benchmark estimators, and more generally calibration estimators, satisfy in particular the following set of properties: 1) they are weighted estimators with the weights as similar to the original sampling weights as possible, given that 2) they estimate perfectly the "known information."

Unfortunately, population totals are usually known only for large cells with units of different types which does not allow us to consider units in a given cell as "similar." Therefore, in practice, calibration estimators are usually applied to non-response adjusted data. The non-response adjustment procedure can be complicated and cause large variances of the estimates. Consequently, it would be very useful to know under what conditions the non-response adjustment procedure can be excluded from the estimation processes without producing essential biases. The goal of the present project is to find these conditions. Results of the investigation are applied to the National Compensation Survey data.

CONCURRENT SESSION III-C:
MODEL-BASED SURVEY ESTIMATION

Potential Applications of Model-Assisted Estimation to Demographic Surveys in the United States
Robert Fay (U.S. Census Bureau)

For decades, the U.S. Census Bureau has conducted a number of large-scale household surveys, including the venerable Current Population Survey. The American Community Survey (ACS) recently became the largest. Until the ACS, samples for these surveys were usually drawn using a mixture of address and sometimes area frames from the previous decennial census, supplemented by frames for new construction. Although census data were reflected in the sampling design, estimation for the ACS and other surveys has been based on complex forms of ratio estimation to independent totals, without further incorporating frame information.

Beginning with Census 2000, the U.S. Census Bureau now maintains a Master Address File (MAF) as a current inventory of housing units. The MAF provides the sampling frame for the ACS and the redesigned versions of the other major demographic surveys. This change to a housing-unit frame presents an opportunity to investigate possible model-assisted approaches to estimation. Model-assisted estimation affords an approach to incorporate information from the frame into the estimation, information that might include data from the previous census or administrative records. Moreover, model-assisted estimation uses auxiliary information in an essentially design-unbiased way, whereas model-based alternatives generally share the risk of introducing

substantial bias. This paper will report on initial results for ACS using data for 36 test counties during 1999–2001. It will also suggest how similar principles could be applied to some of the other demographic surveys.

New Improved Small Area Models

Swamy Paravastu, Tamara Zimmerman, and Edwin Robison (Bureau of Labor Statistics, USA)

In the context of sample surveys, an estimator for a domain (or a subpopulation) such as a geographic area, a socio-demographic group, or an industry group is referred to as "direct" if it is based only on the domain-specific sample data. A domain is regarded as large if the domain sample size is large enough to yield direct estimates of adequate precision; otherwise, the domain is regarded as small. Reliable estimates for small domains can only be produced by moving away from the design-based estimation of conventional direct estimates to indirect estimates derived from correctly specified models. The challenge with the modeling approach lies in the last phrase: How to find correctly specified models? If a model is seriously misspecified, then it can yield inferences that are worse than design-based inferences. One of the findings of this paper is that some of the models that yield design-consistent estimators do not limit the effects of model misspecifications. A statistical model that a model-dependent approach employs may appear to "borrow strength" in making an estimate for a small area from its regressors if the effects of its misspecifications are ignored. The claim that an indirect model-dependent estimator for a small area leads to large gains in efficiency can be incorrect. A correctly specified model may not, in fact, lead to such large gains.

Smoothing Covariance Matrices From Survey Data Using Generalized Design Effects With Application to Small Area Estimation

Avinash Singh and Yong You (Statistics Canada), Ralph Folsom and Akhil Vaish (RTI International, USA)

Typically in modeling for estimating first order parameters such as small area means, one needs to specify the functional form of unknown second order parameters for combining estimating functions with weights determined by the inverse of the covariance matrix as well as for computing precision of estimates. If this form is not too complicated, then one can either estimate the second order parameters directly under an empirical Bayes framework or deal with them under a hierarchical Bayes framework. However, in modeling with survey data, the functional form of the second order parameters or the covariance matrix of estimating functions corresponding to small areas may be quite complex. As a result, the estimated covariance matrix is generally too unstable in the sense that it could be highly ill-conditioned, and may attach unreasonably too high or too small weights to some estimating functions. In such situations, it may be preferable to first smooth the covariance matrix via a separate model, using perhaps the past data, and then treat them as known. This was done in the well known Fay-Herriot (FH) approach to SAE. A fully Bayesian solution would be to use an inverse Wishart-type prior on the design-based covariance matrix; see e.g., Otto and Bell (1995).

In this paper, we propose a new but simple type of modeling for smoothing the design-based covariance matrix of estimating functions applicable to the FH method and to the EFGL (Estimating Function-based Gaussian Likelihood) method of Singh and Folsom (2001) and Singh, Folsom, and Vaish (2002, 2003), the latter is useful when considering a unit-level nonlinear generalization of FH modeling. The proposed method is based on the idea of smoothing the generalized design effects (g-deffs) which is somewhat similar to that used in Rao-Scott (1981) relating the covariance matrix under a complex design with that under simple random sampling. Here, we consider, instead, covariance matrices when the design is ignorable or not for the model, and the g-deffs are defined by the eigen values (λ_i's) of the matrix $\Sigma_s^{-1}\Sigma_c$ where Σ_s is the simple covariance matrix of area-specific estimating functions under the simplified assumption of ignorable design, and Σ_c is the complex design-based covariance matrix.

In the above Σ_s is assumed to be stable, and if necessary, estimating functions may be collapsed for this purpose, while Σ_c is expected to be unstable due to small or large λ_i for those areas with small sample sizes. Smoothing of eigen values λ_i's, in turn, smoothes the covariance matrix Σ_c in view of the matrix result on a pair of real symmetric matrices with at least one of them being positive definite (cf: Rao, 1973, p 41) which states that there exists a nonsingular matrix S such that $\Sigma_c = S\Lambda S' = \sum_i \lambda_i S_i S_i'$, and $\Sigma_s = SS' = \sum_i S_i S_i'$, where $\Lambda = diag(\lambda)$. It follows that by smoothing λ_i's such that no value is too small or too high (i.e., by removing the ill-conditioning), and noting the stability of Σ_s, the desired matrix Σ_c becomes stable.

We will consider different models for smoothing λ_i's and compare empirically via a simulation study for ignorable and nonignorable clustered sample designs for linear mixed models. Among others, we consider a simple smoothing model $\lambda_1 = \overline{\lambda}$ for all areas which is similar to the Rao-Scott first order correction to the chi square statistic for categorical survey data, as well as its extension by assuming constant eigen values over a small number of subgroups where each subgroup consists of eigen values with similar magnitudes. Another possibility for repeated surveys such as the Canadian Labor Force Survey might be to average g-deffs over time. Application to CLFS for estimating unemployment rates for subprovincial areas will also be presented for illustration.

IRT and Latent Variable Modeling for Surveys With Complex Sampling Design With a Longitudinal Context: The Case of the National Longitudinal Survey of Children and Youth in Canada
André Cyr and Alexander Davies (Statistics Canada)

A unique study of Canadians from birth to adulthood, the National Longitudinal Survey of Children and Youth (NLSCY) provides a single source of data for the examination of child development in context, including the diverse life paths of normal development. The NLSCY is designed to follow a representative sample of Canadian children from 0 to 25 years of age, with data collection occurring at two-year intervals. The current sample of NLSCY children is large enough for analysis by cohorts, sub-populations and provinces. Starting In 1994, the first year of data collection, the sample of about 20,000 children then aged 0 to 11, were selected and information about their home and school environment is being collected.

Item response theory (IRT) offers many advantages to researchers who need to quantify children's reading and writing abilities, and for this reason, IRT methods have been adopted in Statistics Canada's National Longitudinal Survey for Children and Youth. IRT methods have a long history in the field of psychometrics, and provide a model-based method for characterising both test items and subject abilities, and for generating predictions of individual abilities. For the most part, IRT methods implicitly assume i.i.d. observations, so that the application of these methods to complex surveys raises a number of issues. Questions arise as to the use or otherwise of the sample weights, appropriate methods for predicting individual abilities (for inclusion on public use datasets), and appropriate methods for estimating parameters and their variances.

The presentation will outline the many steps taken by Statistics Canada to apply psychometric techniques to derive appropriately measured items parameters in the construction of assessment tools and to derive relevant ability scores for children to support longitudinal analysis. This will be an opportunity to showcase some of the research efforts and findings from the last few years of the issues of psychometric testing in a complex setting of longitudinal data and surveys with complex designs. Specific empirical results based on NLSCY data will be provided, including: (1) the potential for biases due to ignoring survey weights; (2) the impact of survey design on variances of parameter estimates; (3) biases in the distribution of ability predictors, and the dependence of this bias on test length. Issues requiring further research will be identified.

CONCURRENT SESSION IV-A
APPLICATIONS OF SURVEY AND STATISTICAL METHODS

Data Collection From Alaskan Cruise Ships
Nelson Andrews, Don Anderson, and Elizabeth Kim (U.S. Environmental Protection Agency)

As the cruise ship industry has expanded over the past several years, the concern over potential impacts on water quality from this industry has grown. Cruise ships are literally floating cities; individual vessels of 3,000 passengers and crew are not uncommon. These vessels typically visit a number of ports, offer many services to their passengers, and operate in ocean and coastal waters. There is significant concern about the environmental impacts of wastewater from these ships, especially in the pristine waters of Alaska. Consequently, Congress provided authority to EPA to develop standards for the discharge from large cruise ships operating in the waters around Alaska. In developing these standards, EPA has initiated two major data collection efforts, which are the focus of this presentation. First, during the summer of 2004, EPA conducted its first-ever comprehensive field sampling program on large cruise ships operating in Alaska. This field sampling was a million dollar-plus effort that required extensive coordination with cruise ship personnel and their off-site managers of four large ships; contractors collecting, processing and analyzing samples; the Alaska Department of Environmental Conservation; and the U.S. Coast Guard. This field sampling will provide data about pollutant discharges and the effectiveness of different treatment technologies. Second, EPA intends to collect financial and technical information through a mail survey that will be sent to the firms that own and operate cruise ships in the Alaskan waters. Ultimately, EPA will select determine whether new standards are needed based upon economic achievability, implementation issues, cost-effectiveness, operational feasibility, and projected environmental benefits.

Innovative Data Collection Methods in Smallpox Program Evaluation
Brian Evans, Brian Burke, and Paul Levy (RTI International, USA)

As part of the national response to the threat of bioterrorist attacks regarding the smallpox disease, Smallpox Response Teams were formed in late 2002 to provide critical services in the event of a smallpox attack. The National Immunization Program (NIP) within the Centers for Disease Control and Prevention (CDC) asked health care workers (HCWs) and other critical personnel to volunteer to receive the vaccine against smallpox. CDC-NIP established a voluntary Smallpox Vaccine Program (SVP) for the HCWs. This program was administered through

56 state and county health departments that were working with hospitals to identify and inoculate health care workers who comprised the Smallpox Response Teams. However, owing to the decentralized nature of the program, there was considerable variation in the administration of the program among the hospitals, health departments, and other state agencies.

CDC-NIP developed the ***Evaluation of Non-Participants in the Smallpox Vaccination Program*** as an opportunity to evaluate the success of the SVP, learn about factors related to how the vaccination program was administered and the reasons why HCWs chose not to receive the vaccination. RTI International (RTI) was contracted by CDC-NIP to conduct this evaluation. The program evaluation itself consisted of a questionnaire administered to HCWs who decided not to receive the vaccination, asking specifically about factors that influenced their decision not to be vaccinated.

However, these data only provided one point of view about the SVP, that of the HCW. In order to supplement the potentially subjective view of the SVP by the HCWs, RTI and CDC-NIP discussed and agreed upon an innovative approach to obtain a potentially more objective perspective about the vaccination program and how it was administered. An additional questionnaire was developed for the Smallpox Vaccination Coordinators at each hospital and health department that participated in the program evaluation. Among other things, this questionnaire asked the coordinator to provide information on how the program was administered, the employer's compensation policies, and barriers encountered during the SVP administration. This approach of administering an additional questionnaire has allowed CDC-NIP to analyze data from two different perspectives to gain a better understanding of how the SVP was administered and why some HCWs chose not to get vaccinated.

The HCW questionnaire provides data about personal reasons for not receiving the vaccine, while the coordinator questionnaire provides more objective data about SVP administration. These data from the coordinator questionnaire provide valuable insight in determining reasons for non-response in the SVP. Additionally, by obtaining data from the smallpox coordinators, aggregate level analysis by hospital and health departments can be performed by linking these data to the interview information provided by the HCW to give an overall picture of how the SVP was administered and why so many HCWs decided not to be vaccinated. Patterns of consistencies or disconnects in opinions about the SVP between the smallpox coordinator and those employed at the hospital or health department about the promotion and administration of the SVP could also be identified. These data also provide insight about how a vaccination program of this magnitude could be carried out in the future, including how CDC might better prepare for future adult vaccination programs.

RTI administered the coordinator questionnaire to 113 smallpox coordinators, 110 of whom completed the questionnaire for a 97 percent response rate. In this paper we analyze questions about promotion of the SVP, compensation to HCWs, and potential barriers or reasons for not being vaccinated. Analysis of these questions in both the respondent and the coordinator questionnaires show how the administration of the coordinator questionnaire provided enhanced insight into reasons for non-participation in the Smallpox Vaccination Program.

Statistical Assessment of the Glare Issue—Human and Natural Elements

Eun-Ha Choi and Santokh Singh (National Highway Traffic Safety Administration, USA)

Visibility is one of the basic requirements for safe driving; any type of obscurity in driver's vision can interfere with the driving task and impose a threat to the roadway safety. Glare is known to be one of the significant factors causing vision obstruction. Although it has been a concern of drivers since oil lamps were substituted by electric headlights, the introduction of high intensity discharge lights has made glare one of the major issues of the roadway safety. One would, however, agree that humans are not the only ones to be blamed for this menace. Often times, nature is responsible, too; bright sunlight can cause vision obstruction that may contribute to crash occurrence.

In this paper, the glare issue and several of its aspects are investigated using statistical methods. Descriptive statistical and configuration frequency analyses are employed in preliminary investigation, while Principal components and Information measures provide a deeper insight into the glare issue. NHTSA crash databases and Omnibus Survey Data 2002 are used in these analyses to assess the overall glare issue. The results can provide guidelines for devising and effectively implementing the glare countermeasures to help the vision-obstructed driver, whatever may be the source of glare.

Value of Physical Change in Farm Inventories: How Does One Reconcile Annual and Subannual Estimates?

Andrew Baldwin (Statistics Canada)

There is a substantial difference in the calculation of the value of physical change of farm inventories in the annual benchmark estimates for GDP by industry, and in the same calculation on a quarterly basis in the quarterly estimates of GDP by expenditure category. In the first case, a single annual price is used to calculate the value of physical change in inventories, in the second case, different prices are used to measure the VPC for each quarter.

The same issues affect the calculation of value of physical change in inventories in volume terms, since an annual base year price can be calculated either directly from end-of-year values, or as some weighted average of quarterly prices, the two calculations not generally providing the same result.

This longstanding issue has become more important recently, now that the monthly indicator for animal production again reflects the value of physical change in inventories. The differences between the two methods are not trivial, and for an agricultural province like Saskatchewan or New Brunswick, the discrepancy between the two is often more important in value than many other individual components of GDP.

The advantages and disadvantages of the two alternatives are considered for both crop and livestock production.

SESSION IV-B

ADVANCES IN FRAME DEVELOPMENT

Rethinking Statistics Canada's Business Register

Eric Rancourt, Hélène Bérard, and Stuart Pursey (Statistics Canada)

In the early 80s, major efforts were made at Statistics Canada to establish a central frame (commonly known today as the Business Register–BR) that could be used by most if not all economic surveys. Indeed, over the years, surveys have gradually been hooked up to the BR and it has become the backbone of the statistical programs of Statistics Canada's Business and Trade Statistics Field. The methodology of the BR was designed to simplify and organise as homogeneously as possible the complex legal and operating structures of businesses. However, since its inception, users of the BR have acquired in-depth knowledge of the structure of businesses. Further, with two decades of economic changes taking place, combined with more uniquely differing surveys that have been joining, and with a processing environment that has naturally aged, the time has come to re-think the purposes and structure of the BR. This paper presents Statistics Canada's current context of business and economic survey-taking and discusses the issues raised to adapt the BR concepts and structure to this new reality.

The Application of Dual System Principles to Estimate the Number of Missing Frame Elements

Howard Bradsher-Fredrick (Energy Information Administration, USA)

The Energy Information Administration (EIA) spends significant resources evaluating and maintaining the survey frames it uses for conducting its large number of surveys. Evaluating survey frames can involve many factors, including direct comparisons to existing survey frames used by other federal agencies and organizations.

One such approach is to directly compare the frame presently employed by EIA with a frame attempting to contain some or all of the same frame units. Such a comparison has several benefits:

1. Effectively updating EIA frames based upon the information available from the alternate frame.
2. Effectively updating the frame of the cooperating agency based upon EIA's frame information.
3. Providing sufficient information so that estimates can be made regarding the number of units missing from both frames.

Significant literature has developed regarding the last benefit (#3) stated. Much of this literature appears to be devoted to discussing the application of dual system principles to specific cases. In its efforts to effectively evaluate and maintain its survey frames, EIA has also employed dual system estimation techniques.

One such recent effort was to compare a listing of renewable electric plants compiled by the National Renewable Energy Laboratory (NREL) with a similar frame employed by EIA. Based upon the number of units common to both frames, or in the EIA frame but not the NREL frame, or in the NREL frame but not the EIA frame, an estimate of the number of units in neither frame could be determined. Moreover, the total frames could be stratified by fuel type and similar analyses could be conducted to estimate the number of units missing from both

frames for each individual fuel type. This paper will present the methodology, computations and results of the application of the dual system principles to the NREL/EIA frame comparison.

Modeling Which Farms Are Not Covered by a Census List Using an Area-Frame Survey

Ted Chang (University of Virginia, USA) and Phillip Kott (National Agricultural Statistics Service, USA)

The National Agricultural Statistics Service (NASS) maintains a list of pontential US farms called "the Census Mailing List." We report on efforts to model the probability of a farm not being on the List in 2002. The analyses were based on data from the NASS area-frame sample and treated that frame as complete. The predictive covariates involved total sales, type of farm, acreage, operator characteristics (gender, Hispanic status, race, and whether the primary occupation of the principal operator was farming), number (if any) of equine on the farm, and the Area-Frame-Survey stratum.

The important predictors were determined to be total sales, farm type, operator characteristics (especially race), and Area-Frame-Survey stratum. Not all farm-type indicators produced significant deviations from the overall predicted probability of a farm being not being on the List.

Several potential link functions were initially investigated, including the logit, probit, log-log, and complementary log-log. Although survey statisticians have been partial to the logit, the probit is popular among economists. Log-log and complementary log-log models have been shown to be particularly effective when almost all or almost none of the population (e.g., all farms) have an attribute (are not on the List).

Predictions of the numbers of farms with certain characteristics were similar using each of the link functions and consistent with area-sample estimates. This led us to use survey-weighted logistic regression in subsequent analyzes. A modification of the stepwise regression algorithm was employed for variable selection. Standard errors were estimated primarily using design based linearization methods.

Various problems for further research into statistical methodology are suggested as a result of our investigation.

TECHNICAL DEMONSTRATIONS

UTE Tools to Test the Usability of Web Sites

Janice Nall (General Services Administration, USA)

The Usability Testing Environment (UTE) is a tool that assists quantitative usability evaluation of Web sites and Web applications. Most of the tools and methods previously dedicated to usability evaluations (including heuristic evaluations, cognitive walk-throughs, and think-aloud evaluations) have been extremely subjective and prone to a potentially destructive phenomenon known as the "evaluator effect." As an automated usability testing environment, UTE provides assistance in all phases of Website usability testing. UTE also automates the capture of usability data in detail not possible by a human observer.

The Usability Testing Environment consists of two tightly-integrated applications. The first is the UTE Manager which helps a tester set up test scenarios (tasks) as well as survey and demographic questions. The UTE Manager also compiles the test results and produces customized reports and summary data which can be used as quantitative measures of usability observations and recommendations. The second UTE application is the UTE Runner. The UTE Runner presents the test participants with the test scenarios (tasks) as well as any demographic and survey questions. In addition, the UTE Runner tracks the actions of the subject throughout the test including clicks, keystrokes, and scrolling.

Using the Usability Testing Environment reduces the time required to compile observations and generate recommendations from a few weeks to a few days—or less. In addition, testers will have actual data to reinforce all observations and recommendations. UTE also requires more thoughtful consideration of what to test during test creation ensuring more useful results.

InFocus and JAWS for Accessibility Testing

Lawrence Malakhof (U.S. Census Bureau)

Since June 2001, Federal regulations legally require that U. S. government Web sites and other software provide comparable access to the information for all users. Computer users who have visual and or hearing disabilities are entitled to have the same access as users who do not currently have any disabilities.

Accessibility is not the same as usability. Accessibility guidelines do have several checkpoints that address usability, such as a logical tab order, dividing large information blocks into more manageable groups, and using the clearest and simplest language appropriate. Even if the application complies with the law, it still may not be

usable, as the Census Bureau's Usability Lab has found in other testing. Both usability and accessibility testing need to be done.

There are a number of software applications that evaluate software for accessibility. At the U.S. Census Bureau we use InSight/InFocus. This tool flags potential accessibility violations, which are checked with the Job Access with Speech (JAWS) screen reader. JAWS is the leading screen reader for MS-Windows applications.

Both software tools will be demonstrated at FCSM this year. FCSM attendees may provide the link to their Web sites for an accessibility violations listing. As time permits, the JAWS screen reader will be used to identify potential usability problems. Handouts will be available to provide guidance on available resources.

Q-Bank System
Kristen Miller (National Center for Health Statistics, USA)

The Cognitive Methods Staff (CMS) at the National Center for Health Statistics (NCHS) developed a database of survey questions evaluated in the Questionaire Design Research Laboratory. The database catalogues the questions and links each question to pretest findings. Questions are searchable not only by *content* or *subject matter* (e.g. asthma questions, cancer questions, demographics), but also by *question type* (e.g. objective characteristics, behavioral reports, attitudes), *response category type* (e.g. yes/no, open-ended, quantity), and *response error type* (e.g. problems with terms, recall problems). Additionally, a statistical tool has been developed to perform basic *analyses of question characteristics* and *identified types of response error*. It is a functional, attractive, user-friendly application.

Plans are underway to expand "Q-BANK." The end product is envisioned to be an online interagency database of tested survey questions and research results, maintained at NCHS and provided on a subscription basis to participating federal agencies.

Innovative Web-Based Documentation System Designed for the National Survey on Drug Use and Health
Inga Allred, (RTI International, USA)

Abstract not available.

Blaise IS (Computer-Assisted Interviewing Software for the Web)
Jim O'Reilly (Westat, USA)

Recently Westat has implemented Web surveys using the Blaise system for survey instruments with large numbers of question items, complex routing, and other features beyond those found in most Web surveys. One study is a multi-modal (CATI and Web) household interview that collects nationally representative data on the public's need for, access to, and use of cancer-related information. Another study is a distributed CAPI application where data are collected and entered in a clinical setting. We will demonstrate these applications and summarize our experiences, including:

- Instrument development issues
- Willingness of household respondents to complete a lengthy survey
- Special considerations for Web surveys targeted for data collection staff
- Web survey performance for large surveys
- Server-side architecture and management issues.

Metadata Software for NYCHANES
Debra Reed-Gillette (Centers for Disease Control and Prevention, USA)

The Centers for Disease Control and Prevention (CDC) operates the National Health and Nutrition Examination Survey (NHANES) for the purpose of collecting data to assess disease and health risk behavior across the non-institutionalized population of the United States. The National Health and Nutrition Examination Survey (NHANES) is conducted both with extensive interviews in a participant's household and an extensive physical examination in specially equipped mobile examination centers. CDC and the DHNES program worked with the City of New York to utilize the HANES program to conduct locally-defined population surveys in a pre-defined community. The DHNES program developed a survey implementation system using survey metadata to drive data collection, data import, data export, data transformation, and analysis. The challenges of the metadata design were myriad, including the support of national and community surveys, multi-hierarchical questionnaires, sophisticated hierarchical item attributes, gender/age text substitutions, complex item flow, questionnaire and item versioning, and definition/inclusion of standards.

We will demonstrate the HANES metadata driven instruments that were based on metadata dynamic interpretation for data collection and data delivery. We will also demonstrate our prototype metadata editor for survey authoring, and automated data export.

Web-Based Factorial Survey and Statistical Analysis Software

Sam Addala (e4Xchange Corporation, USA), Matthew Hogben (Centers for Disease Control and Prevention, USA), and Vin Addala (e4Xchange Corporation, USA)

Virtually anyone seeking information from a given group of people faces this dilemma: as one asks for more information, potential participants become more reluctant to provide all the information. Such reluctance is as likely to be on the basis of time and trouble as on grounds of sensitivity. Factorial surveys, however, permit researchers to sample different combinations of information from different respondents, thus allowing different respondents to contribute data to a subset of the full survey. To date, existing factorial survey software does not allow a researcher naVve to programming to design a survey and sampling protocol, and gather data (via internet) under the same umbrella.

e4Xchange Corporation, a software development organization and a NASA industry partner, has developed an intuitive and user friendly software solution to address the end-to-end needs of researchers conducting factorial and non-factorial surveys for the Centers for Disease Control and Prevention. e4Xchange has completed the Phase I proof of concept and has moved on to Phase II, full scale development. Proof of concept entailed developing factorial algorithms such that the survey designer could specify dimensions and levels of dimensions, which would then be sorted into all possible combinations and randomly presented.

Phase II is being built around four functions: design, Web publishing, response, and analysis. Design is the cardinal function, with the user able to develop "static" survey questions with varied response options, as well as factorial questions with varying numbers of factors and levels of factors. The designer simply writes a story (vignette) in a text box and names factors directly within that vignette. The designer then defines and names levels of factors, earmarks invalid combinations for partial factorial designs, and lets the program do the rest. Eligibility criteria and block randomization, if needed, can be built into the design, and the designer can specify or randomize the order of questions (or even just some of them), as well as denote skip patterns. Notably, the software is unique in that the user need not be a computer programmer to build complex factorial surveys.

Respondents log on, anonymously or confidentially, and either as a convenience sample or according to user-designed sampling protocols. The output is readable in a spreadsheet format, and therefore via common statistical programs (e.g., SPSS or SAS). Potential statistical analysis of output ranges from descriptive statistics to experimental frameworks from the General Linear Model (e.g., analysis of variance/regression, curve estimation). The program has Web publication functions: when to publish surveys on the Web, for how long, and for how many respondents. Survey access can be Web-based, e-mail, or hard copy.

In sum, the software is designed to accommodate the needs of data-gathering and statistical analysis requirements of government, educational, and private industry.

ACS Automated Cell Suppression System

Gordon Sande (Sande and Associates, Canada)

The ACS Suite of software builds upon the technology proven in the Statistics Canada CONFID system to provide a comprehensive solution to the problem of developing and auditing of cell suppression patterns for the protection of the confidentiality of respondents to establishment-based surveys. The new implementation uses an easily understood command language which can be used by end users and which they can use as documentation of all their processing steps. The methods have been extended to include additional heuristics to determine complementary suppressions, an analysis to explain the complements for each sensitive cell or the sensitive cells using each complement, self consistent table completions and an extended analysis of groupings of cells which are sensitive. Other extensions include range publication using variable base rounding and analysis for price indices. The self contained system starts from a micro data set, and related metadata, of establishment responses and returns a control file specifying the publication status of the cells in the publication. The aggregation of establishments to enterprises, the allowance for sampled units in an otherwise census of units and the identification of sensitive cells is done in a tabulation component. The working tables follow the structure used in the cell suppression analysis with preparation of tables for publication left to the main survey automation. The suppression with its alternate heuristics and the audit with optional table completion and adjustment to additivity are separate components. Utilities provide for specification of unique subject matter requirements of the survey. The entire process is scripted and rerunnable with full provenance of all intermediate steps maintained.

The utilities are designed to allow execution of many production scenarios. Early releases can be marked as already published with the automation completing the publication pattern. Forced publication or suppression

of selected cells can be indicated to allow for unique subject matter requirements. A publication pattern can be established for an almost final micro data set and then rerun quickly when the final micro data set is available. This allows both a full confidentiality review and a low publication delay. Automation would permit alternate patterns to be prepared for professional review. A periodic publication may apply the ongoing publication pattern to the current data for auditing only. Historical continuity can be aided by treating the previous publication pattern as a starting point to have additional suppressions or releases to reflect changes in data sensitivities.

Variable Base Rounding. Variable base rounding is a technique intended to reduce the intrusiveness of cell suppression by a stylized range publication. It uses the many available sources of ambiguity in a combined fashion. The higher relative error of low level cells is explicitly combined with the lower relative error of high level aggregates. The data is complete with its ambiguity explicitly indicated by the stylized range. A separate component implements this new technique.

Price Index Auditing. Many business activities are measured twice with the ratio of the measures also of interest. In business statistics of manufacturing or refining this is volume, price and revenue data, for labor statistics this is hours, wages and payroll data and in agricultural statistics this is crop land, yield and harvest data. Volume, price and revenue data is often published as volume and price data as revenues would be redundant. A separate specialized audit component was developed to address the unique issues of publishing such highly related data.

Controlled Rounding. Similar techniques and architecture have been used for the problem of controlled randomized rounding. Multi-way and hierarchical tables have controlled rounding with user indication of where adjustments may be placed if required. The adjustments may optionally be sign restricted. The additional adjustments are indicted when they are required.

Survey Data Modeling With Mplus
Tihomir Asparouhov (Muthen & Muthen, USA)

Mplus is a comprehensive modeling program that integrates random effect, factor, and latent class analysis in both cross-sectional and longitudinal settings and for both single-level and multi-level data. We present new Mplus features for multivariate analysis with complex survey data. The general modeling framework underlying Mplus is described. This framework allows unparalleled modeling flexibility with a variety of observed dependent variable types, both continuous and categorical (mixture) latent variables, and several ways of handling complex survey data. The framework includes analyses as varied as SEM, IRT, growth modeling, latent class analysis, and multi-level modeling. Mplus capabilities will also be featured in computer demonstrations.

CONCURRENT SESSION V-A
VARIANCE ESTIMATION

A Study of the Properties of a Bootstrap Variance Estimator Under Sampling Without Replacement
Lenka Mach and Jean Dumais (Statistics Canada) and Lauriane Robidou (University of Social Sciences, France)

This paper evaluates the performance of the bootstrap variance estimation in the case of a two-stage survey design with high sampling fractions at the first-stage. Multi-stage sampling is commonly used by statistical agencies, for example for household or education surveys. Typically, samples are selected without replacement (WOR) but the standard bootstrap method assumes that the PSUs are selected with replacement (WR) or the first-stage sampling fractions are negligible. Thus variance overestimation is suspected when high fractions of PSUs are sampled WOR at the first stage. Modified versions of the bootstrap for designs with WOR sampling have been proposed, but these are restricted to single-stage sampling of clusters or two-stage designs with equal-probability sampling of PSUs.

We use a simulation study, based on the data from the Statistics Canada Youth in Transition Survey (YITS) of 15-year-olds, to illustrate the extent of the bias as well as the stability of the bootstrap variance estimator. At the first stage of YITS, a stratified sample of schools is selected and, at the second stage, students within selected schools are surveyed. Due to requirements for precise estimates in small sub-populations, PSU sampling rates in some strata are as high as 60 percent. So far, we have studied the properties of the bootstrap variance estimator for simple statistics, totals and means, for small, medium and large strata of schools. Our results suggest that the first-stage sampling rate is not the only factor determining the bias; the second stage sampling rate seems to play a role as well.

An Empirical Investigation Into the Effects of Replicate Reweighting on Variance Estimates for the Annual Capital Expenditures Survey

Katherine Thompson (U.S. Census Bureau)

The Annual Capital Expenditures Survey (ACES) collects data about the nature and level of capital expenditures in non-farm businesses operating within the United States. Respondents report capital expenditures for the calendar year in all subsidiaries and divisions for all operations within the United States. The ACES universe contains two sub-populations: employer companies and non-employer companies. Different forms are mailed to sample units depending on whether they are employer (ACE-1) companies or non-employer (ACE-2) companies. New ACE-1 and ACE-2 samples are selected each year, both with stratified SRS-WOR designs. The ACE-1 sample comprises approximately seventy-five percent of the ACES sample (roughly 45,000 companies selected per year for ACE-1, and 15,000 selected per year for ACE-2). The ACE-1 sample is highly stratified, with companies within industry classification stratified into five separate size strata. Especially in the large-company size strata, ACE-1 sampling rates are often non-negligible; sampling rates in approximately one-fifth of all strata are greater than 0.20.

The ACES estimation procedure is fairly straightforward. Sample weights are adjusted for unit non-response, and these non-response adjusted weights are used for subsequent estimation. The ACES variance estimation procedure is equally straightforward: ACES uses a "simple" delete-a-group jackknife variance estimator to produce estimates of levels and uses Taylor linearization methods to produce variances of change estimates. "Simple" refers to the replicate weights, which are constructed from the full-sample non-response adjusted weights; currently, ACES does not replicate their non-response adjustment procedure. This variance estimator is computationally quick and easy and does not require excessive overhead in terms of computer storage.

The decision to use simple delete-a-group jackknife variance estimation for ACES was based on research presented in Thompson, Sigman, and Goodwin (2003), which used both empirical data (one year) and simulated data. Somewhat surprisingly, that study found few—if any—statistical benefits in replicating the non-response adjustment procedure with those data. This finding was not only counterintuitive, but was also inconsistent with several other published studies. We hypothesized that our unexpected results could be caused by a variety of factors, such as high unit response in large company strata, non-negligible sampling fractions in large company strata (the results cited above assume negligible sampling fractions), or an inappropriate choice of variance estimator (with a highly stratified survey, the stratified jackknife might be more appropriate).

This paper examines the effects of performing replicating the non-response adjustment procedures on empirical variance estimates of capital expenditures statistics using three years of ACES data. We consider two jackknife variance estimators (the delete-a-group jackknife and the stratified jackknife) while considering all of the factors listed above.

Investigations of BRR Variance Estimation for the Survey of Residential Alterations and Repairs (SORAR)

Laura Ozcoskun and Katherine Thompson (U.S. Census Bureau) and Quatracia Williams (ICS-RIO, Brazil)

The U.S. Census Bureau publishes quarterly estimates of expenditures for the improvement and repairs of residential property. These estimates are obtained from the Consumer Expenditures Survey (CE) and the Survey of Residential Alterations and Repairs (SORAR). The SORAR sample is subsampled from the CE, which is a probability sample of households designed to represent the total U.S. noninstitutional population.

SORAR uses Balanced Repeated Replication (BRR) to estimate the variance of its quarterly estimates. The SORAR weight adjustment procedure consists of four separate adjustments to the initial sample weights. The presently-used BRR implementation does not replicate this weight-adjustment procedure: it uses a "shortcut" approach, multiplying the full sample adjusted weights by the appropriate replicate factor. Many other studies have found that this shortcut approach underestimates the variance, often by a large amount, especially when the magnitude of the adjustments to the original weights are fairly large (Canty and Davison 1999, Yung and Rao 2000). The BRR variance estimates constructed using this shortcut approach may be considerable underestimates, since they do not directly account for nonreponse adjustment variability, the ratio adjustment variability, or the units on property adjustment variability. Because SORAR has a very low monthly response rate (approximately 15 percent), there is some concern that underestimating the variance could result in greatly exaggerated estimates of precision.

In this paper, we investigate whether all—or some—components of the SORAR weight adjustment procedure should be replicated for variance estimation, comparing the effects of full replicate reweighting (all stages of the weight adjustment procedure replicated), partial replicate reweighting (two of the four stages of the weight adjustment procedure replicated), and no replicate reweighting (the current method) using historic SORAR data. We consider the effects on the size of variance estimate, practical impact in terms of c.v.s, and computation time in our evaluation.

CONCURRENT SESSION V-B:
CONFIDENTIALITY

Evolutionary Cell Selection for Interval Publication of Sensitive Cells to Protect Confidentiality
Timothy Li and Steve Cohen (Bureau of Labor Statistics, USA)

With public demand for more information from government statistical publications, publishing sensitive cells previously suppressed now in predefined, fixed intervals is an attractive alternative. However the published fixed interval bounds may reveal more information to the reader and increase the confidentiality exposure risks that had not existed before. Simply replacing previously suppressed cells with pre-defined, fixed intervals does not provide the level of confidentiality protection provided under the complete suppression method. Therefore a completely new set of cells has to be selected for interval publication purpose, and more cells may have to be published in intervals as a consequence. This demands a new method, in reference to the methods used for complementary suppression cell selection methods, to select a set of additional protection cells (PC) to also be published in fixed intervals. In this research we propose an iterative "selection-improvement" algorithm, which improves cell selection upon each previous suppression pattern until all primary cells are sufficiently protected. During each iteration, additional loops of PCs are added to the previous suppression pattern, in such a way that primary cells were at risk from previous iteration will be protected in the current generation. Constrained optimization is used to detect risky primary cells. Local loops are selected to minimize information loss among feasible solutions. This algorithm always guarantees a solution. Though it is yet to be shown an overall minimization of loss of information has been achieved, a localized minimization is essential to this heuristic. In addition, this method builds upon current cell protection activity at survey program offices by adding additional improvement steps, therefore it is easier to implement in practice than other methods. We apply this method to the publication table consisting of five sectors in a U.S. state from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages to show its application in practice.

Federal Statistical Confidentiality and Business Data: Challenges and Continuing Issues
William Seltzer (Fordham University, USA) and Margo Anderson (University of Wisconsin—Milwaukee, USA)

The roots of the modern concept of federal statistical confidentiality can be traced directly back to the late 19th Century efforts of such statisticians as Carroll Wright, founder of the Bureau of Labor Statistics, to ensure full and accurate responses by businesses to statistical inquiries. Wright and others argued that such confidentiality guarantees were needed to ensure that the providers of enterprise and establishment data could be confident that the statistical agencies could not be forced to share their responses with others, such as regulatory or tax authorities, congressional investigators, prying journalists, and competitors, who might use this information to the detriment of the data provider.

It is widely recognized in the federal statistics community and by knowledgeable members of the business community that the logic of this position is as true today as when it was first articulated over a century ago. (Indeed, the principle of statistical confidentiality based on analogous reasoning was extended to information provided by persons beginning with President Taft's proclamation issued in connection with the 1910 decennial census, a topic which the authors have discussed elsewhere.) Nevertheless, over the years, the principle of statistical confidentiality with respect to information provided by businesses in statistical inquiries has been repeatedly challenged by other executive branch departments, independent regulatory agencies, the courts, Congress, and members of the public, with quite varied results.

This paper will review this history of challenges and the responses of the concerned statistical agencies, the federal statistical system as a whole, including the office of the chief statistician in OMB (and its predecessors), executive department and independent non-statistical agencies, the courts, and Congress as well as representatives of the business community. While one instance—involving the protection that could be afforded to file copies of Census Bureau questionnaires maintained by a respondent (the St. Regis case) culminating as it did in an adverse decision by the Supreme Court and an act of Congress over-turning that decision and reaffirming the Census Bureau position—is reasonably well known, others are apparently unremembered. For example, the Federal Trade Commission, the initiator of the effort to obtain the data provided to the Census Bureau by the St. Regis Paper Co., also initiated efforts to obtain enterprise-level data collected under a pledge of statistical confidentiality from other federal statistical programs, including that of the U.S. Bureau of the Mines.

The paper will draw on the published record and the results of recent archival research by the authors to examine the historical record, and based on that, to discuss (a) any long-term trends suggested and (b) implications for maintaining and strengthening the confidentiality protections for establishment- and enterprise-level business data provided to federal agencies for statistical purposes.

Privacy Principles and Data Sharing: Implications of CIPSEA for Economic Surveys Respondents
Alfred Tuttle and Diane Willimack (U.S. Census Bureau)

Historically, sharing data among federal statistical agencies has been prohibited by tight legal restrictions emphasizing the privacy and confidentiality of information provided by individuals and business entities. Certainly the safekeeping of confidential information submitted in good faith by the public deserves our full attention. However, it results in an overall data collection system that is burdensome for business respondents in particular, especially for large companies that are continually selected for many different government surveys. In addition, some core items must be collected independently by multiple agencies conducting establishment surveys, in order to aid descriptive classification of establishments for summarizing or analyzing the data. This too adds to business survey respondent burden.

To alleviate some of the burden arising from redundant data requests, Congress enacted the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002, which permits the sharing of business data among three designated statistical agencies—the U.S. Census Bureau, the Bureau of Economic Analysis (BEA), and the Bureau of Labor Statistics (BLS)—under carefully prescribed conditions designed to uphold the confidentiality of information submitted by the public. While CIPSEA does not explicitly require informed consent from respondents at the time of data collection (except regarding data collected for non-statistical uses), the Census Bureau is considering an informed consent policy that would require survey programs to make respondents aware of "planned uses" of the data, one of which may include data-sharing. Since data-sharing is permitted only for business data, such a policy would need to be administered differently for economic programs than it is for demographic surveys, due to CIPSEA.

This paper describes research into known and anticipated concerns of business survey respondents vis-B-vis data-sharing permitted by CIPSEA. This paper will also evaluate materials from selected Census Bureau economic survey programs relative to these concerns, along with criteria suggested by the Bureau's Privacy Principles and other legal guidance. Our findings will be a springboard for further research to develop and test respondent materials for administering privacy and informed consent principles in Census Bureau economic programs, given the possibility that respondent-reported data may be shared with CIPSEA-designated agencies.

CONCURRENT SESSION V-C:
GEOSPATIAL PERSPECTIVE AND ANALYSIS

The Geospatial Perception and Its Impact on the Content and Processes of a Multi-Source Data Collection
Oliva Blum (Israel Central Bureau of Statistics)

The introduction of the geospatial perception to the statistical information world has stimulated the imagination of people as of what information can and should be produced, and has enhanced the abilities to do so. It has enabled the integration of information, the specification of complex spatial models describing population characteristics and behavior, the analysis of volumes, areas and distances in space, over-time etc. The implementation of this perception in the statistical production results in changed and usually improved processes. The integration of heterogeneous information-sources and data processing tasks are inherent to the infrastructure and their quality depends on the integrity and timeliness of the system. Sampling and small area estimation enjoy the homogeneous stratification enabled by territorial divisions, drawn along the boundaries of spatial phenomena. The improved spatial orientation simplifies fieldwork operations, and the statistical products are of improved quality and diversity.

The geospatial perception has been recruited and developed as part of the census methodology in Israel. The qualitative geospatial infrastructure has been the enabler of the transition from a traditional census to an integrated one. This census is based on the central population register (CPR) that has to be corrected for census purposes. Two sources of information are used in the correction process: administrative files and field surveys. The administrative files correct the CPR directly, through editing and imputation processes, and their end product is an integrated administrative file (IAF). However, the field surveys supply the information needed for small area estimation, and their results are translated to weights assigned to each individual record in the IAF.

Data collection from administrative sources is supported by spatial record linkage and data selection processes. The differential ability to anchor information and consequently integrate data from different administrative sources is a crucial selection criterion for the source of information to be used. The components of the quality evaluation are of geospatial nature and include: the extent and accuracy of the possible geo-references, the resolution of the geo-reference, and the coverage completeness of the file with regard to the census area (census space). Despite the careful process, not all records in the IAF are anchored to the smallest spatial unit, and some data elements are geo-referenced to aggregated spatial entities.

Data collection in the field surveys are guided by the need to evaluate the over and under-coverage of the IAF in each estimation units. The definition of the sampled population is a geospatial process in both surveys. In the undercoverage evaluation survey (UES), area-cells are delineated according to the characteristics of the geo-referenced CPR population, and according to the boundaries of the statistical area. The sampling process itself is also using geospatial analysis based on the statistical area borders, the tendency to be under or over-covered in the administrative files, the proximity between sampled area cells, and the fieldwork-organization requirements, related to the data collection routes and to the interviewers workloads. The undercoverage estimation is based on the comparison between the population found in the sampled area cells and the population in the IAF that is spatially anchored to the same statistical area. In the overcoverage evaluation survey (OES), the sample unit is an individual record in the IAF. However, in order to save resources, this sample overlaps, fully or partially, the records anchored to the sampled area-cells of the UES. Only those people whose administrative records carry a residential address in the sampled area cells, but have not been interviewed in the UES, are to be found in the OES. Only those who are found to be out of the statistical area are labeled as overcoverage.

Many census processes are inherently geospatial. Therefore, in the long run a register-based approach has to be applied in the maintenance of the infrastructure. All spatial entities have to be uniquely identified and covered and all events have to be recorded as close as possible to their occurrence. In the interim, the incomplete geographic infrastructure may obstruct the generation of the census results, and therefore, special data collection processes are required. An example of a locality without addresses in presented.

The Geospatial Distribution of Employment
Sheryl Konigsberg, David Talan, and Richard Clayton (Bureau of Labor Statistics, USA)

The advent of powerful computing capabilities and mapping software now allows more sophisticated analysis of new and existing problems through the visual display of information. The center point of these new features is the ability to provide pinpoint locations for geographic features. These locations are defined by precise latitude and longitude coordinates, called geocodes. In any geocoding system involving businesses, the key is to have accurate physical location addresses.

This article profiles the growing interest for geocoded data, examples of existing applications using labor market information, and early efforts to obtain and use geocodes for the Bureau of Labor Statistics (BLS) business establishment list in the Quarterly Census of Employment and Wages (QCEW) program. The QCEW program serves as a census and constitutes the most complete set of monthly employment and quarterly wage information by six-digit industry at the national, state, Consolidated Metropolitan Statistical Area (CMSA), Metropolitan Statistical Area (MSA), and county levels.

Traditionally, the QCEW program has provided economic data at the national, state, and county level. Data from the QCEW serve as an important input to many BLS programs as well as other federal and state programs. These data are used as a benchmark for the Current Employment Statistics and Occupational Employment Statistics. The QCEW is also used by the Bureau of Economic Analysis for GDP and personal income estimates.

Geocoded data presents labor market information in a new dimension. Demands for more local data are providing an incentive to tabulate data for cities, towns, and even smaller areas. With the availability of geocoded data, BLS will be able to develop lower levels of aggregations, including cities, postal zip codes, census tract, and census block. These data can then be mapped using GIS system, and when combined with a range of other economic, social, geographic, demographic, and other data, can create a rich tapestry of information that allows users to analyze information quickly and examine data in new and exiting ways.

Cities of Immigrants: Intraurban Mobility Patterns of Mexican Immigrants in Gateway Cities
Pamela Rogers (University of Texas at Austin, USA)

Census 2000 identified the impact of immigration upon a number of emerging gateway cities. Little is known, however, about the spatial mobility and work-seeking experiences of recent migrants and immigrants in these metropolitan centers. Specifically, where they live at the outset, what governs their search behavior for employment, how far they mobilize social capital in the first few months, and about their medium- to long-term employment trajectories. Similarly little is known about their parallel residential trajectories, as they move from being sharers or renters upon arrival to later tenure housing arrangements as they become settled in the medium- to long-term. Whereas considerable information is available about immigrants arriving to the United States and then traveling to and from Illinois or California, or between metropolitan cities such as Chicago and Los Angeles, little is known about the intraurban migratory patterns of immigrants in Texas. This study addresses these remaining questions about how immigrants settle and adjust within metropolitan areas using 5 percent metropolitan level data from the Public-Use Metropolitan Sample (PUMS) and aggregate-level analysis.

Residential and Demographic Patterns of Immigrants In Texas at a Glance: Focus on Mexican Foreign-Born Immigrants

Yann-Yann Shieh (American Institutes for Research, USA) and Pamela Rogers (University of Texas at Austin, USA)

A recent review of immigration literature reveals a growing discussion of the importance of "emerging gateway cities" (Suro and Singer 2002; Singer 2003; Frey 1995; Frey 2003). Much of the research taking place is based on the unprecedented growth of foreign-born populations in places which had not experienced any such notable growth before the 1990s. Recent policy studies from the Brookings Institution and the Center for Immigration Studies focus on the impact of immigration upon urban population growth, urban and suburban development, housing, city services, and commuting. Within this setting, this research studies the residential and demographic changes taking place in Texas which have attracted an unprecedented number of primarily Mexican immigrants. Using the 2000 5 percent Public-Use Microdata Sample (PUMS), both descriptive analyses as well as logistic regression analysis include a wide range of metropolitan-level data derived from the person and housing records. The findings suggest that the number of persons in the household, overall household income in 1999, household structure, as well as language isolation were significantly related to homeownership. Gender, English ability, citizenship status, as well as personal total income in 1999 were significantly related to high-school education attainment. Similarly, gender, English ability, citizenship status, as well as a person's high school education attainment were significantly related to person's total income. This research contributes to the broader studies of Mexican immigrant research by applying empirical approaches to one of the first studies of immigrants in Texas.

CONCURRENT SESSION VI-A

STATISTICAL MEASURES

Total Factor Productivity Computed and Evaluated Using Multi-Step Perturbation

Baoline Chen (Bureau of Economic Analysis, USA) and Peter Zadrozny (Bureau of Labor Statistics, USA)

We describe and illustrate a method for computing and evaluating total factor productivity (TFP). First, we describe using the multi-step perturbation (MSP) method to compute TFP based on any k+1-times differentiable production function. We illustrate the method for Cobb-Douglas (CD), constant elasticity of substitution (CES), and tiered CES (TCES) production functions. Thus, we describe and illustrate computing TFP for far more general production functions than the CD production function, which is the basis of the usual Solow-residual computation of TFP. Second, simultaneously, we describe and illustrate a method for choosing the empirically most valid production function and implied TFP. The MSP method computes optimal inputs, hence, input residuals, the differences between observed and optimal inputs. Optimal inputs maximize output, for a given production function, input prices, and input costs. An information criterion (IC) has a log-likelihood term, computed as the determinant of the sample covariance matrix of input residuals, and a parameter-penalty term, which increases with the number of estimated parameters. Usually, the production function—the model in this case—which implies the lowest IC is considered the best or empirically most valid production function among those being considered. In effect, the usual Solow residual sets all input residuals identically to zero, which results in estimated parameters and implied TFP having no degrees of freedom and no statistical reliability. By contrast, we use the MSP method to compute TFP based on the "best" CD, CES, or TCES model with the lowest IC and compare it with the usual Solow-residual TFP. We do this using a sample of data on capital, labor, energy, materials, and services (KLEMS) inputs, from 1949-2001, obtained from the Bureau of Labor Statistics.

New Data on Business Employment Dynamics

Akbar Sadeghi, Richard Clayton, and David Talan (Bureau of Labor Statistics, USA)

One of the most watched economic indicators in the United States is the monthly change in nonfarm payroll employment released by the Bureau of Labor Statistics. Changes in this indicator can affect stock market movements and interest rate decisions considerably. The monthly change in nonfarm payroll employment gives the *net* change in the number of jobs over a particular month—the overall change, given that some establishments have opened, some have expanded, some have contracted, and some have closed. This article presents new BLS data on quarterly *gross* job gains and losses, documents their magnitude, and examines the historical time series having to do with these statistics from 1992 to 2003.

The new BLS measures of gross job gains and gross job losses afford a more thorough understanding of the employment decisions of the millions of business establishments in the U.S. economy. Examining establishment-level employment changes aids in analyzing both the large gross job flows that underlie the substantially smaller net employment changes and the establishment-level employment dynamics across various stages of the business cycle.

The article begins with definitions of *gross job gains* and *gross job losses*. Following this introductory section are a description of the source data used by the Bureau to generate estimates of quarterly gross job gains and gross job losses and an explanation of the methodology employed for longitudinally linking establishment records. The heart of the article is the presentation of the new BLS business employment dynamics data series, together with an analysis of the levels and movements of gross job gains and gross job losses during the past 10 years. Special attention is given to technical issues such as the seasonal adjustment of gross job gains and gross job losses, how the business employment dynamics data compare with other BLS establishment-based employment series, and establishment openings and closings relative to births and deaths. The article concludes with a summary of ongoing work and planned future enhancements to the gross job gains and gross job loss statistics at the Bureau.

Evaluating Estimates of Labor Demand and Turnover
Charlotte Mueller and John Wohlford (Bureau of Labor Statistics, USA)

The Bureau of Labor Statistics has developed a new establishment survey that is producing nationally representative data on labor demand and turnover in the United States. In adding to the federal statistical framework, these data are shedding new light on the U.S. labor market and economy in general. This paper begins by briefly describing the Job Openings and Labor Turnover Survey (JOLTS). The paper then describes some of the efforts to evaluate the quality of the estimates, specifically highlighting comparisons of the JOLTS data series with other recognized and comparable series. These comparisons help validate the long-run trends seen in the new JOLTS series, and statistical correlations confirm what past research has suggested. The job openings estimates have a strong inverse relationship with the unemployment rate as depicted by the Beveridge Curve, and the job openings series mirrors the trend movements of the Conference Board's Help-Wanted Advertising Index. Hires estimates are strongly correlated with employment change, and trends in separations match the movements of a smaller turnover report from the Bureau of National Affairs. As quits dominates the movement of total separations, it has been compared with consumer confidence measures and shows similar trends over time. All of these comparisons lend validity to the JOLTS data series, presenting new opportunities for the series to contribute to a wide range of labor market research and analysis.

Measurement of Reliance on Social Security Benefits
T. Lynn Fisher (Social Security Administration, USA)

Over thirty years ago, the Social Security Administration began producing a popular statistical series, The *Income of the Population 55 or Older* based on data from the Census Bureau's Current Population Survey. The series continues to provide statistics on the receipt of income from various sources, income distribution, aggregate income, and poverty. The most widely cited statistic is the proportion of total income that elderly beneficiary aged units receive from Social Security benefits.

Financial planners have often used the metaphor of the three-legged stool needing all three legs of savings, pensions, and Social Security when discussing the adequacy of retirement income. The elderly receiving all of their income from Social Security are considered economically vulnerable, and the percentage of the elderly population completely dependent on Social Security for income is considered an important indicator of the economic well being of the elderly. As reported in *Income of the Population 55 or Older*, the percentage of elderly beneficiary aged units receiving all of their income from Social Security benefits remained stable around 13 to 15 percent from 1978 to 1992, but then proceeded to climb unabated to 22 percent in 2002

This rise in complete dependence on Social Security from 1992 to 2002 is puzzling considering that the official poverty rate for the elderly decreased from 12.9 percent to 10.4 percent (U.S. Census Bureau 2004b) over the same time period. In addition, greater numbers of people were participating in a booming financial market, which would suggest that a greater proportion of the elderly would have asset income than in previous decades, implying a lower rate of complete dependence on Social Security benefits.

These observations point to two broad research questions: what may explain the reported increase in dependence and how best to measure such dependence. This paper begins a consideration of the second of these questions by conducting a series of "what ifs" for 1996 to explore the various factors used to measure the reliance on Social Security benefits. What if 90 percent or more reliance were considered? What if the unit of observation were a person or a family? What if the Survey of Income and Program Participation (SIPP) were used instead of the March Supplement to the Current Population Survey (CPS)? What if administrative data on Social Security benefits, Supplemental Security Income (SSI), and earnings were substituted into the surveys for reported data? In addition, tradeoffs involved with these alternatives are discussed.

CONCURRENT SESSION VI-B
EVALUATING THE RESPONSE PROCESS

Participation in the National Health Interview Survey: Exploring Reasons for Reluctance Using Contact History Process Data

Nancy Bates and Andrea Piani (U.S. Census Bureau)

In 2002-2003, the Census Bureau designed an automated contact history data collection system known as the Contact History Instrument or CHI. The CHI was developed as a stand-alone Blaise instrument designed to work across all Census Bureau demographic Computer Assisted Personal Interview (CAPI) surveys as part of the case management system. Interviewers use the CHI to record the number of contact attempts, mode, date and time of attempt as well as details behind various interim outcomes (e.g., reasons for refusals and strategies attempted).

Using CHI data from the 2005 National Health Interview Survey CHI we explore the reasons why some households are reluctant to participate in the interview process. We base the analysis on 10 weeks of CHI data representing over 45,000 visit attempts to 15,000 households. We investigate the extent of reluctance, what the most frequently cited reasons are, and whether these vary by characteristics such as household composition and other auxiliary variables such as region, urbanicity, or mode of contact. We will also report how patterns of reluctance may change as the number of contacts increases. Finally, we attempt to uncover if there are underlying dimensions that represent common elements of reluctance and whether some reasons are more highly correlated with the decision to refuse the survey. In closing we offer recommendations how CHI data can be used as a feedback mechanism for improving field productivity and understanding the reasons people participate in federal surveys.

The Effect of Interviewer Strategies on Contact and Cooperation Rates in the National Health Interview Survey

Barbara Stussman, James Dahlhamer, and Catherine Simile (National Center for Health Statistics, USA)

The decline in completion rates for Federal surveys has been a major concern for Federal statistical agencies in the past decade (Atrostic, et. al., 2001). Nonresponse rates for the National Health Interview Survey (NHIS) have more than doubled over this period. For the interviewer, the challenge is two-fold: she must first make contact with sample household member(s) and then gain their participation in the survey. Improving upon, or even maintaining, current response rates depends, in part, on the efficiency with which interviewers negotiate this two-stage process. Interviewers employ a variety of strategies, such as checking with neighbors and leaving promotional materials, to enhance their chances of making contact and gaining cooperation, yet few systematic investigations have looked at this topic.

Detailed contact history data provide a unique opportunity to explore the relationships between interviewer strategies and survey response. A Contact History Instrument (CHI) was implemented in 2004 to capture critical information on each NHIS contact and contact attempt. Interviewers use CHI, for example, to record information on access impediments and strategies employed to induce contact and cooperation. As such, CHI provides a record of interventions used to combat nonresponse in the NHIS. This paper explores the effect of various interviewer strategies, such as leaving notes, promotional materials, “no one home” letters, checking with neighbors, and others, on making contact and gaining subsequent cooperation. More specifically, we address three research questions. (1) What strategies are interviewers using and in what combination? (2) How do interviewers shift strategies to suit specific field situations? (3) How do strategies used on prior attempts relate to contact and cooperation outcomes on subsequent attempts? By exploring the use of interviewer strategies in the NHIS, we can increase the efficiency of their use in the field and enhance the development of additional strategies for reducing nonresponse in surveys.

Socio-Demographic Study of Telephone Survey Nonrespondents

Timothy Triplett and Natalie Abi-Habib (Urban Institute, USA)

Response rates for telephone surveys in the United States are declining. The impact of nonresponse on estimates derived from a telephone survey depends on the extent to which nonrespondents differ from respondents on survey items of interest. Whether survey estimates are unbiased depends on the assumption that nonrespondents are missing at random (MAR). When nonrespondents differ from respondents the MAR assumption does not hold and nonresponse error may be present in the sample.

Little empirical evidence exists to evaluate the MAR assumption because any information about nonrespondents must come from administrative records or other auxiliary sources. The current research examines the differences between respondents and nonrespondents to the 2002 National Survey of America’s Families (NSAF). Auxiliary information about nonrespondents comes from Census 2000 at the block group level. Block groups

are the smallest level of census data available publicly and are estimated to encompass between 250 and 550 housing units. Census data are used to create typologies that characterize neighborhoods (defined at the block group level) as metropolitan or non-metropolitan, urban or rural, predominantly of one racial/ethnic grouping or another, and as poor or wealthy. Census typologies are linked to NSAF households at the block group level in order to compare the neighborhood typology traits of NSAF respondent and nonrespondent households. Learning the extent to which the neighborhood characteristics of nonrespondents differ from those of respondents sheds light on the appropriateness of the assumption that nonrespondents are missing at random and the potential for nonresponse error in the sample.

Are Two Feet in the Door Better Than One? Using Process Data to Examine Interviewer Effort and Nonresponse Bias

Kevin Wang, Jeremy Aldworth, Rodney Baxter, and Joe Murphy (RTI International, USA)

Survey organizations employ numerous tactics to reduce the potential for bias due to unit nonresponse. After data collection, nonresponse and population weighting adjustments are often utilized in order to reduce potential bias from nonresponse and undercoverage. Prior to and during data collection, interviewers are trained, and sometimes re-trained in techniques for gaining cooperation from reluctant respondents. In addition, interviewers attempt to reduce the potential for bias from nonresponse by revisiting or recontacting sampled households, and attempt to persuade reluctant respondents to cooperate. In some cases, supervisors may assign highly skilled interviewers to persuade the most reluctant respondents to cooperate.

The amount of effort that interviewers can apply to achieve high response rates is limited by time and resource constraints. That "effort" may be regarded as a resource to be allocated in an efficient manner in order to reduce the potential for bias from nonresponse while simultaneously maintaining a dedicated and productive field staff. Indiscriminate application of extensive efforts to gain cooperation may not reduce the potential of bias from nonresponse in an effective manner. In principle, effort could be differentially applied to cases based on the potential for reducing bias from nonresponse (either by lowering the nonresponse rate or targeting cases that might be especially different from current respondents). In order to manage field staff in this way, survey organizations require information on the likelihood of successfully completing an interview, especially with regard to factors that are potentially under management control such as how field interviewers manage their time and organize their assignments. In addition, these efforts should be informed by an understanding of how interviewers allocate their efforts and whether the level of effort is related to survey measures of interest.

In this paper, we use administrative call record data from the National Survey on Drug Use and Health (NSDUH) to examine these issues. The NSDUH is an in-person cross-sectional study of nearly 180,000 households and 67,500 persons conducted in all 50 states and the District of Columbia each year. The survey is administered primarily through an audio computer-assisted self-interview (ACASI) via a laptop computer; a short household screening interview collecting demographic data for use in respondent selection is administered by interviewers. The NSDUH is designed to measure the prevalence and correlates of drug use in the United States population age 12 and older. Respondents completing the one-hour interview receive a $30 cash incentive.

CONCURRENT SESSION VI-C:

LINKING ADMINISTRATIVE AND SURVEY DATA

The Estimation of the Italian Households' Financial Assets

Ivan Faiella, Leandro D'Aurizio, Stefano Iezzi, and Andrea Neri (Bank of Italy, Italy)

The Italian households' financial assets are evaluated mainly through the Financial Accounts (quarterly revised and updated) and by the sample Survey on Household Income and Wealth, carried out every two years. The Bank of Italy is responsible for both sources. The sample estimates are uniformly below those coming from the Financial Accounts, even after harmonising all the definitions in the two sources. The difference between the value declared by the interviewee and the corresponding actual value is a major source of the aforementioned gap. Such a problem can stem either from the interviewee's unwillingness to disclose the ownership of an asset or from a wrong declared value, generally lower than the actual one. For each asset the first type of error is defined as *"under-reporting on the possession"*, whereas the second one is labelled as *"under-reporting on the amount"*. The paper sets forth a method for improving the sample estimates of the Italian households' financial assets by exploiting a sample survey of customers of a leading Italian commercial bank. The main feature of this survey is that asset values are available both from the interview and as registered in the bank's customer database: these data make it possible to model underreporting behaviors and to extrapolate the results to the Bank of Italy's survey. Data are corrected at the household level; at first a simulation process attributes the possession of an asset to households who do not declare it, but have a high probability of possessing it (correction for the under-reporting on the possession), subsequently an asset amount is imputed for these households.

At the same time the declared asset values are increased to adjust for the under-reporting on the amount. The whole process produces a substantial increase of the sample estimates.

Evaluation of CPS Tax Simulation Using Administrative IRS Data
Amy O'Hara (U.S. Census Bureau)

A new methodology was implemented to simulate federal and state taxes for the Current Population Survey (CPS). Person-level survey responses were compiled into filing units to estimate taxes. The Census Bureau uses the tax estimates to compute after-tax income, and releases many of the variables on the public use dataset. Administrative data from the Internal Revenue Service (IRS) was obtained and used to evaluate the modeling effort. The results of linking the CPS records with individual income tax records were used to gauge how well return classifications and the number of exemptions were assigned. These results will allow better modeling of filing incidence in the future.

Evaluating Respondents' Reporting of Social Security Income in SIPP Using Administrative Data
Lydia Scoon-Rogers (U.S. Census Bureau)

This paper takes a look at how well respondents report Social Security income in the Survey of Income and Program Participation (SIPP). This investigation was done by linking person records from SIPP to a special research file from the Social Security Administration (SSA) called the Payment History Update System (PHUS). The SSA's PHUS data file is an *actual record of what was paid monthly* to Social Security beneficiaries and reflects all adjustments for any double checks sent, death, or for other administrative accounting irregularities. This SSA file is considered "closer to truth" than SSA's Master Beneficiary Record that has been used in most research that has assessed reported social security income. This latter SSA file only indicates what beneficiaries were *supposed to receive* in any given month, not accounting for any adjustments.

Comparing reported Social Security amounts of SSA beneficiaries in SIPP to the actual amounts available from the SSA administrative data has uncovered several types of respondent error. This paper looks at the extent to which improper accounting of the Supplemental Medical Insurance (SMI), a component of Social Security income, has affected reporting and to what extent other types of error, such as rounding and neglecting to report cost-of-living adjustments, appear to be contributing to SIPP's overall underreporting of Social Security income for seniors.

CONCURRENT SESSION VII-A
ADMINISTRATIVE DATA AND DATA QUALITY

Reporting Data Quality When Survey and Administrative Data Are Combined
Julie Trépanier, Claude Julien, and John Kovar (Statistics Canada)

At Statistics Canada, business surveys, both annual and monthly, are using administrative data at an ever increasing rate. The administrative data come from either payroll deduction accounts, income tax reports or Goods and Services Tax reports collected by the Canada Revenue Agency. These data are not only used to build and maintain the frame, the Business Register, or to assist imputation of survey data, they are now used to replace completely or partially subpopulations that would have traditionally been surveyed. The primary aim is to reduce response burden and survey cost. The survey data are either replaced directly, or modelled using the administrative data, relying on a strong correlation between the administrative data and the survey data. As such, a typical survey estimate is now based on both survey and administrative information. The source of the data may vary by unit and even by variable.

Traditionally, data quality indicators reported by surveys have been the sampling variance, coverage error, nonresponse rate and imputation rate. Are these indicators still relevant in a context where administrative data are used? In fact, it is not unusual for some surveys to produce estimates that are based largely on the administrative source and thus reporting virtually no sampling error. In 2004 a task force was set up to look into these issues and to recommend a strategy on how to report data quality in the context where survey and administrative data are combined. An attempt is made to produce a measure of total error associated with the reported estimates. This paper presents the progress of the work accomplished by the task force.

Developing Error Prone Profiles Using Administrative Data With a Control Group

Pedro Saavedra and Hoke Wilson (ORC Macro, USA)

The practice of error-prone profiling to select applicants for various entitlement programs has made use of a variety of prediction techniques, from multiple linear regression to logistic regression to regression trees. In most programs, data exists where a random sample of applicants or participants have been audited, and a resulting outcome can yield a disposition indicating whether the original data provided by the program participant was in error. In the Student Financial Aid (SFA) system, one only has access to administrative records indicating if a person was selected to be audited or not, if their data changed, and if he or she was paid. But students change their data spontaneously, or drop out for reasons unrelated to the error or the audit. This paper describes various methodologies that have been explored to derive profiles from administrative data under these constraints. The research design involves examining changes made by a group of SFA applicants randomly selected for verification, and changes made by a set of SFA applicants randomly selected to be immune from verification. The procedures involved the derivation of a pool of potential predictors (selection criteria) and an equation or model to combine the selection criteria. Through the years prioritizing the criteria through an effectiveness index, a regression tree method based on the phi correlation coefficient and the combination of multiple regression equations with dummy variables as predictors have been used. Most recently C&RT has been used to generate dichotomous variables for the pool of predictors.

Using the Canadian Address Register in the Labour Force Survey: Implementation, Results, and Lessons Learned

Claude Turmelle, Jean-François Rodrigue, and Gavin Thompson (Statistics Canada)

Since the Canadian Labour Force Survey (LFS) was first introduced, it has always had a multi-stage design for which the primary source of addresses used to form the sampling frame were actual listing exercises performed directly in the field for each selected Primary Sample Unit (PSU or cluster). But since the last LFS redesign, which was implemented in 2004/2005, a new source of addresses is now being used; the Canadian Address Register (AR). The AR is a list of over 12.5 million addresses (almost 90 percent of the dwellings in Canada) that was built from administrative sources and then improved and extended using the Canadian Census of population.

In this paper, we will briefly describe the LFS design, the AR and its link to the National Geographic Database and the impact it has on its quality. We will also briefly describe the AR update mechanism.

Also, even if having access to a list frame was opening the door to a one stage design, for many reasons it has been decided that LFS would once again keep its two stage design. Because of this decision, the idea was then to find the best way to use the AR in a two stage design context. Different possibilities were explored and finally one was proposed. Basically, it consists of dividing the population of PSUs into three different groups, depending on the expected quality of the AR. For group 1, where the AR is expected to be of excellent quality, we use it as a frame and directly select the sample of dwellings from it. For group 2, where the AR is expected to be good, but not excellent, we use it as a preliminary list to aid the field staff in preparing the list of dwellings. Finally, for group 3, where the AR is absent or expected to be of poor quality, we don't use it at all, and the field listing continues to be done as in the past.

To test and refine the proposed methodology, a field test was performed in fall 2003 and we will describe this test here as well as present some of the results.

Based on the positive results of the field test, the proposed methodology was accepted and was going to be implemented. Results of the field test were also used to refine the criteria for inclusion in groups 1, 2 and 3, described above. The implementation will be described here, including how the three groups were formed, what kind of feedback we first received from the field and the adjustments that had to be done because of it.

Some early results on gains and losses, in terms of under-coverage and over-coverage, associated with the use of the AR will also be shown.

Finally, some potential future projects, for example using the AR to go to a one stage design, will be discussed.

Methodological Challenges in Analyzing Patient-Reported Outcomes in a Clinical Trial

Elizabeth Hahn and David Cella (Evanston Northwestern Healthcare, USA) and G. Alastair Glendenning (Novartis Pharmaceuticals Corporation, USA)

The analysis of patient-reported outcomes (PRO) data in an international clinical trial presents methodological, statistical and interpretive challenges. Specific strategies are needed regarding the psychometric measurement properties of self-report instruments, cross-cultural measurement equivalence, definitions of clinical significance, missing data, longitudinal modeling, and descriptions of clinically interpretable results. Development and implementation of these strategies is illustrated using the 16-country International Randomized IFN

vs. STI571 (IRIS) Study of 1,106 newly diagnosed patients with chronic phase chronic myeloid leukemia. The primary endpoint was the duration of progression-free survival; PROs were secondary endpoints. Crossover to the other treatment was permitted because of intolerance or lack of efficacy. The Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) was completed as a measure of health-related quality of life at baseline, months 1–6, 9, 12, 18, and 24 in the patient's preferred language. The methodological issues and specific strategies developed to address them are summarized. An item response theory (IRT) measurement model was used to evaluate psychometrics, including cross-cultural comparability (eight languages), and to aid in interpretation of treatment differences. A mixed effects model was chosen for the longitudinal analyses, with a pattern-mixture technique to adjust for nonignorable missing data. Crossover effects were added as a time-dependent covariate. To better understand the meaning of the PRO scores, a clinically significant treatment effect was prespecified, and a modified forest plot was used to summarize IRT responses. 1,049 patients (95 percent) participated in the assessments. The patterns of dropout and change were quite different for the treatment arms. This study presented major methodological challenges to PRO data analysis, all of which were addressed using state-of-the-art modeling techniques. The analysis plan and results may be useful for statisticians, researchers and clinicians who analyze and interpret PROs.

CONCURRENT SESSION VII-B:

ANALYZING AND IMPROVING RESPONSE

Earlier Versus Later Respondent Assessments on Customer Surveys: Is There a Difference?

Howard Bradsher-Fredrick and William Weinig (Energy Information Administration, USA)

Federal agencies are increasingly using customer surveys as one measure among several to evaluate their services and products. This is due in part to the more widespread use of performance measures and strategic planning as mechanisms to improve the quality of their products and effectiveness in providing services.

Toward that end, the Energy Information Administration (EIA) administered customer surveys on a sample of customers twice in 2004. In both situations the issue of how much to pursue non-respondents arose, leading to the question of whether they differ from respondents. Using later respondents as a proxy for non-respondents, we compared the later respondents to earlier respondents to determine whether these were significant differences in responses.

The first study was conducted in March 2004; the products evaluated were the annual publications, *The Annual Energy Outlook* (AEO) and *The International Energy Outlook* (IEO). The population surveyed was the registrants to the National Energy Modeling System Conference. This is an annual conference in which the latest edition of *The Annual Energy Outlook* is discussed. All of the non-EIA pre-registrants to the conference were surveyed; this amounted to a first mailing of 450. The e-mail system was employed as the vehicle for distribution.

The second study was conducted in September 2004; the product evaluated was the *Short Term Energy Outlook* (STEO). The target population was the non-EIA users of that product. A random sample of 500 was drawn from the 4,000 users registered on the Listserv. Again, e-mail was employed as the vehicle for distribution.

In both studies, a total of 4 mailings were administered with a total response of 133 (33.2 percent) and 122 (25.9 percent), respectively. The number of questions employed in evaluating each of the 3 products varied, but was approximately 14 for each. The evaluation questions employed a five-point Likert Scale.

For each study, the survey data were stratified into two strata: respondents to the first mailing and respondents to the second through fourth mailings. These then became operationally defined as the earlier respondents and later respondents, respectively. This paper will present a non-parametric analysis to determine whether significant differences existed in the responses of the earlier versus later respondents to each question.

For the AEO and IEO, no significant differences were found for the responses to any question. For the STEO, the evaluations showed a generally lower score to the questions for earlier respondents, but only one question was significantly lower at the .05 level. Thus, our findings were not really definitive on the issue of whether or not earlier respondents are different from the later respondents, but differences in the population may play a role in our findings.

Results from Recent Experiments on Improving Response Rates

Ron Fecso (National Science Foundation, USA)

The National Science Foundation conducts a series of surveys of individuals with a bachelor's degree or above. The recently completed set of surveys, including the 200,000 plus sample National Survey of College Graduates, included various experiments aimed at improving response rates in future waves of the panels. Experiments included use of incentives, variations in introductory letters, variations of automated messages, use of a brochure, variations in mode of first contact, and targeted letters for nonresponse follow-up. This paper will present the results of the experiments.

Efficient Estimation of Response Rates When a Small Subsample of Nonrespondents Is Selected for Follow-Up Conversion

Avinash C. Singh (Statistics Canada), Vincent G. Iannacchione, Shijie Chen (RTI International, USA), and Jill A. Dever (Joint Program in Survey Methodology, USA)

For surveys with a nonresponse follow-up, the usual double-sampling estimator offers a way to measure the combined response to the initial survey and the follow-up. However, when cost considerations require that the follow-up sample size be small, the double-sampling estimator may be unstable. In this paper, we use dual-frame estimation techniques to obtain a more precise composite response rate (CRR) which combines the unstable but unbiased double-sampling estimator with the stable but biased estimator of response rate from the main survey after being bias-corrected.

Under a population response model, each unit in the population is assigned a random response indicator of 1 or 0 before the actual sampling takes place. Suppose the parameter of interest is the total number of respondents if the survey were administered to the entire population. In this case, the response rate parameter is the total number of respondents divided by the population size. For approximately unbiased estimation of this population response rate, it is natural to use the design-weighted estimator especially if the sampling design is non-ignorable.

When a survey is subject to low response, a follow-up of initial nonrespondents can provide a better understanding of bias attributable to nonresponse. In this situation, it is desirable to estimate the response rate if all individuals in the entire population who are nonrespondents to the main survey but subject to possible conversion via a follow-up are also included. For this new parameter, the response rate estimator from the main survey will be biased because of under-reporting because it doesn't use the respondents who could have been converted.

To correct for under-reporting bias among the nonrespondents in the main survey, we propose to use a mass imputation of means under a super-population model which models the probability of an individual responding to the follow-up survey conditional on the individual being nonrespondent to the main survey. We use the sampling weight calibration approach to fit this imputation model using respondents from the main and the follow-up surveys. We illustrate the proposed method of estimating CRRs using data from the 1991 Gulf War Veterans Survey.

Tests of Methods to Improve Response to Physician Surveys

David Woodwell and Catharine Burt (National Center for Health Statistics, USA)

Background:

Health surveys targeting physicians historically have had difficulties in obtaining high response rates. Response rates to the National Ambulatory Medical Care Survey (NAMCS) have been declining since 1985 (from ~75–65 percent). In the NAMCS, a sample of non-federal office-based physicians is sent an introductory letter explaining the survey and stating that a Census Bureau Field Representative (FR) would be calling to make an appointment. At the interview, the FR asks the physician some eligibility questions and other questions about the practice. The physician is then asked to complete a one-page form for each of about 30 sampled patient visits to the office during the randomly-assigned reporting week. The Patient Record form (PRF) collects information about the patient such as age, sex, and race and the visit such as diagnosis, diagnostic services and treatments ordered or provided. CDC conducted several studies between 2000 and 2002 to try to gauge the impact of different methods on physician response to the NAMCS. One of the goals of the survey is to keep response above 70 percent at a minimum. The techniques explored sought to raise response rates to that level so that nonresponse bias would be minimized.

Methods:

Samples of physicians were randomly assigned to treatment and control groups in three separate studies to test the effects of the following on NAMCS response: 1) inclusion of a motivational insert with the introductory letter containing Q & A's about the survey and its importance, 2) offering monetary ($50) or token gift ($15 value of candy, flowers, or other token) incentive at the time of survey induction, and 3) amount of content requested (single vs. 2-sided page with twice the content). Response rates were compared between the treatment and control groups for each study separately.

Results:
There were no differences in physician response for the motivational insert, or the incentive groups. The longer Patient Record form was found to negatively impact response by lowering the response rate from 68 to 62 percent.

Conclusions:

Physicians will continue to be a very difficult group to survey in the 21st century. The policy implications include that decisions based on health care delivery estimates from surveys of physicians in the US may continue to be compromised even further. The focus may need to switch from improving response rates to estimating and controlling for nonresponse bias.

CONCURRENT SESSION VII-C:
SURVEY INTEGRATION AND PLANNING

Chair: Polly Phipps (Bureau of Labor Statistics, USA)

Integrating the Canadian Annual Survey of Manufactures Into the Unified Enterprise Survey: Challenges and Issues
Isabelle Marchand, Emmanuel Benhin, and Jose Gaudet (Statistics Canada)

The Annual Survey of Manufactures (ASM) is one of the largest of all the business surveys conducted by Statistics Canada. It is one of the main contributors to the calculation of the Canadian Gross Domestic Product, representing almost 20 percent of the economic production. The ASM covers all establishments belonging to the manufacturing sector and the logging industry. It collects various financial data such as employment, wages, cost of raw materials and supplies used (referred here as input total), total sales of goods manufactures (referred here as output total), and inventories. It also collects detailed commodity data such as the cost of each raw material, the sale of each manufactured product, and the quantity of a manufactured product. As part of a strategic streamlining initiative that aimed to reduce survey costs and resources while maintaining a certain level of data quality, the ASM conducted some major changes for the reference year (RY) 2004. One of the objectives of these changes was to maximize the use of fiscal data. These changes affected many components of the survey process such as the sample design, questionnaire contents, methodology and processing systems.

One of the major changes was to further integrate the ASM into the Unified Enterprise Survey (UES). These two surveys have shared a few components since RY2000. The Unified Enterprise Survey is an annual business survey that serves as the vehicle for producing annual estimates for many industries at a variety of geographic levels. The different annual surveys that are part of the UES are integrated into one platform where a single frame, similar methodology, and processing systems are used. Another important objective of the UES is to reduce response burden of Canadian businesses by increasing the use of administrative data (Gauthier, 2005).

In this paper, we present some gains and limitations of survey integration in a general context. A summary of the overall challenges and issues from the integration of the ASM into the UES is presented. This is followed by detailed discussions on the integration processes and their associated challenges for each of the ASM survey steps. Finally, we provide some advantages and disadvantages of the integration process.

Implementing and Improving Quality Standards at the U.S. Census Bureau

John Bushery and Pamela McGovern (U.S. Census Bureau)

Disseminating quality data is a key issue for federal data collection programs. In October 2001, the Office of Management and Budget (OMB) issued government-wide information quality guidelines that provide policy and procedural guidance to federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information they disseminate.

As the largest statistical agency of the federal government, the U.S. Census Bureau recognizes that high quality data are essential to help improve our nation's social and economic well-being. The Census Bureau's mission is to provide "quality data about the nation's people and economy." To support this mission, in 1998, the Census Bureau initiated a Quality Program. One component of the Quality Program is the Quality Framework, which consists of documents that communicate the Census Bureau's position on quality. These documents include principles, standards, guidelines, current practices, and best practices for the Census Bureau's survey and statistical processes and products. In early 2004, the Census Bureau created a full-time Quality Program Staff. The Quality Program Staff's responsibilities include assisting the program areas with implementing the Quality Standards, and evaluating the implementation of those standards and providing recommendations for improving them.

Recently, the Quality Program Staff took on the assignment of benchmarking the Census Bureau's Quality Standards against those of other statistical organizations, to identify "gaps" in the Quality Standards and recommend corrective actions to fill those gaps.

This paper presents a brief overview of the Census Bureau's Quality Program, an example of the process for evaluating the implementation of the Quality Standards, and a brief discussion of the benchmarking exercise.

Developing a Generic Survey

Marla Smith and Jan Matuszko (U.S. Environmental Protection Agency)

Over the last 30 years, the U.S. Environmental Protection Agency (EPA) has conducted business surveys of numerous industrial sectors. These surveys have collected economic and technical information used to develop regulations for industrial water pollution control. For each industry, EPA has designed unique sampling plans and questionnaires by evaluating factors such as production processes, typical wastewater treatment, and business practices. This development phase is extremely time-consuming. This presentation will describe our efforts to develop a generic screener questionnaire that can be quickly tailored to different industries. These screeners would be used to obtain a broad overview of an industry and to better identify the target population for a more detailed questionnaire. While industries have common elements, the terms used by each can be different, and thus, we will need to confer with other EPA staff and industry representatives to develop questions that will be relevant to multiple industries, while still providing information that can be used in the regulatory process. In addition, the level of sophistication varies from industry to industry, and sometimes within an industry. We will explore possible solutions, such as the feasibility of telephone interviews, rather than mail surveys, because this mode might allow us to better tailor the questions to the respondent, and reduce the rate of nonresponse.

Integrated Redevelopment of the Office of National Statistics' Business Surveys and Business Register Through the Business Surveys Integration Project and Register Re-Engineering Project

Mark Pont, Pam Tate, Paul Smith, and John Perry (Office for National Statistics, United Kingdom)

The present portfolio of business surveys of the Office for National Statistics (ONS) has grown up piecemeal in response to different demands and concerns, at different times, and within different organizations. There have over time been various reviews and redevelopments, most notably the creation of the consolidated Annual Business Survey in 1998, but most of the reviews have been specific to individual existing surveys, and have tended to focus on the detail of existing processes. We still have (depending on exactly how we count them) around 70 different ONS business surveys, using some 800 different survey questionnaires.

ONS is currently modernizing its infrastructure under the moniker of the Statistical Modernisation Programme (SMP). ONS is also moving to common data management systems, and at the same time standardizing its methods and implementing common tools with which to apply them, and systematizing its processes. Through this program, ONS expects to make considerable gains in both the quality of its outputs and the efficiency with which they are produced.

But we can say with fair confidence that further gains can be made, in both quality and efficiency, through rationalization and integration of the wide and disparate range of our existing business surveys. Two major projects are involved in realizing these additional gains: re-engineering the UK's business register (the Inter-Departmental Business Register [IDBR]), and the Business Surveys Integration Project (BSIP).

The BSIP has the broad objective of developing a redesigned and integrated portfolio of business surveys to meet the major requirements for economic statistics. This includes: revising survey methodology in a consistent way that applies current best practice and facilitates optimization of survey design and coherence of outputs; meeting emerging needs such as those for extended and improved outputs at regional level, and for improved detail of the service industries sector; and increasing the efficiency and value for money of the program of business surveys by making the best possible use of administrative data sources.

The success of this project is in turn heavily dependent on the extent to which the re-engineered IDBR supports these objectives, for example in providing a sampling frame with a capability to support a variety of stratification and estimation methods, and a unit structure that supports flexible data collection and can make maximum use of information supplied by businesses at varying levels of detail.

The paper describes the development of the two projects, and the ways in which the objectives of integration of the various surveys, and of consistency and mutual support between the integrated surveys and the business register, have been addressed.

CONCURRENT SESSION VIII-A:

QUALITATIVE EVALUATION

Chair: Alethea Jennings (Energy Information Administration, USA)

Determining Best Practices for Usability Test Methodology: A Comparison of Assessment Methods

Elizabeth Dean, Michael Schwerin, and Kimberly Aspinwall (RTI International, USA)

Usability testing is an important method for examining how well system users understand and efficiently use a system to complete a specific task. This research method has led to improved development, evaluation, and testing of various automated questionnaires, information systems, and Web sites. RTI International recently implemented a large-scale, multi-site usability testing effort to assist in the development of the U.S. Navy's new Web-based performance management system. As part of this testing effort, several methodological variations of usability testing were applied including: subjective and objective usability measures, capture of timing data and problem incidences, and conversational (think-aloud) testing and task-oriented testing. This paper will describe the relative utility of these usability methods and examine the differences in findings that resulted from each. The data analyzed will include behavior coding of usability tests and timing estimates from usability tests, as well as usability survey and focus group results.

Timing Data and Problem Incidences: During the actual usability test component of the evaluation, users completed a series of usability test scenarios. For each scenario, task durations and behavior coded problem incidences were analyzed. Task durations highlighted functional problems as well as the overall burden of multiple usability problems on the user experience. These durations were used to assess the costs in labor hours and in dollars of not making improvements to the system, but did not provide many insights as to how the system should be improved. On the other hand, problem incidences provided specific detail on how users interact with the test system. An examination of the number and types of problems associated with tasks led to detailed recommendations for specific improvements to the system.

Subjective and Objective Usability Data: Both subjective and objective measures of usability were collected in each of three rounds of data collection. Subjective measures were reported in pretest and post-test surveys and in post-test focus group interviews. Objective measures were obtained from the behavior coding and timing analyses of the usability test procedures. The objective usability data provided key insights into the burden associated with the most difficult components of the performance management system. The most burdensome steps included logging in, setting a password, and finding the performance management documents. The subjective data provided rich details on a different aspect of implementing the new performance management system—the effect of the change on Navy culture. The subjective results included three main recommendations: 1) incorporate a paper Quick Reference Guide to assist with navigating the performance management system, 2) ensure that one-on-one personal counseling sessions are maintained with the transition to an automated system, and 3) develop a communication plan so that supervisors can guide their direct reports on the implementation of the system.

Conversational (think-aloud) Testing and Task-Oriented Testing: A portion of the interviews were conducted in a conversational environment, in which the test facilitator sat side-by-side with the user facing the test system. The remaining interviews were conducted in an on-site lab-like environment with the test facilitator in a separate room or cubicle. Varying the type of interview was not observed to affect the participant's results or overall comfort level with the usability test. The conversational approach did uncover more usability errors with the test system.

In summary, we found that using each of these three methodological variations led to additional insights and enhanced the quality of feedback we were able to provide to the Navy's performance management system developers. To the extent possible, we recommend varying methods in future usability test protocols.

Cross-Cultural Issues in Survey Translation: Translation of Meaning and Meaning of Translation
Daniel Geller and Andrey Vinokurov (ORC Macro, USA), and Tamara Martin (U.S. Department of State)

The paper outlines the translation process involved in the ORC Macro's evaluation of the State Department's International Visitors Leadership Program (IVLP) in Russia, Ukraine, Kazakhstan, and Georgia. The IVLP is a long-running program in which foreign professionals and prospective leaders from all over the world have the opportunity to participate in funded short-term visits to the United States to improve their professional practices and career positions, as well as to provide opportunities for them to learn first-hand how democratic institutions and processes, free-market economies, and other values of Western democracy and American society are manifested in professional and daily life. The present paper addresses and adds to the growing literature on translation issues by delineating a step-by-step translation process that could be used as a blueprint for adjusting translation methods to fit a particular cultural context. This process highlights the importance of attending to the theoretical issues in the translation process, outlines specific phases of the translation process, presents the modified de-centering translation technique adapted for the project, describes the types of translation equivalences that were addressed, and discusses contextual factors inherent in the translation process. In the process of addressing such a variety of issues, the paper underscores that the meaning and the purpose of the translation process is to provide a qualitative approach for the instrument development that maps contexts of people's lives, documents emic-etic aspects of cross-cultural research, and fosters collaborations with all stakeholders of the research project.

Using Affective Imagery to Understand the Quality of Survey Response
Patricia Gwartney (University of Oregon, USA)

In the past few years, the author and her graduate students have explored affective imagery as a tool for understanding the quality of telephone survey response. Affective imagery refers to the meanings or mental images that certain stimuli evoke in survey respondents. We evoke affective images with open-ended word association questions of the form: *"What is the first thought or image that comes to mind when you hear the word _____?"* As stimuli, we have used words and terms consistent with each survey's topic, including "global warming", "inequality," "organic food," "cloning," "blood," and even "survey," "U.S. census," and "political poll." We then code respondents' narrative, open-ended answers for patterns, and map those patterns with both direct and indirect indicators of response behavior.

Our broad patterns of initial findings indicate that patterns of affective images vary significantly with:

- respondents' opinions about the importance and impact of scientific survey research;
- non-intrusive measures of the quality of respondent behavior, such as interview length, number of words per open-ended question, counts of "don't know" and "refuse" answers, counts of "other-specify" answers, and whether the respondent gave an initial refusal before completing the interview; and
- respondents' socio-demographic characteristics.

These findings suggest that affective imagery yields robust indicators of respondent task dedication and task alienation in completing surveys. Moreover, we have discovered that the quality of survey response for respondents providing **visual images** is significantly higher than those who do not. In one paper, we found that 40 percent of respondents could not supply a visual image, even when probed, and that percentage did not vary across the three stimuli. No matter what their specific answer, the "visualizers" compared to the non-visualizers:

1. ascribed higher levels of importance to survey research in general, and
2. produced less item nonresponse, more words per open-ended question, and fewer minutes per completed interview.

Does Qualitative Evaluation Measure Up?
Kathleen O'Connor (National Center for Health Statistics, USA)

This presentation will describe the role and impact of qualitative observation in refining the content and design of the 2001 National Survey of Children with Special Health Care Needs (NS-CSHCN) in preparation for a repeat survey in 2005–2006. This periodic cross-sectional survey is sponsored by the Maternal and Child Health Bureau (MCHB) of the Health Resources and Services Administration (HRSA), and is conducted by the Centers for

Disease Control and Prevention's (CDC) National Center for Health Statistics (NCHS) via the State and Local Area Integrated Telephone Survey (SLAITS), with contractor assistance. The purpose of the survey is to provide estimates of the number and characteristics of children with special health care needs (CSHCN) in each state and to characterize their health and functional status, the types of services they need and use, and the shortcomings in the system of care. It is the only data source that provides uniform CSHCN data at the state level for program evaluation, planning, and policy development; allows comparisons to be made across states and nationally; and provides a comprehensive array of data on child and family experiences with the health care system.

Although the 2001 NS-CSHCN provided useful comprehensive data, several issues emerged that mandated changes to the earlier instrument and have prompted the initiation of critical methodological research. These issues include:

- improving the description of the population and the impact of the special need,
- examining the validity of the current screener across subpopulations, and
- including children considered to be "at-risk" of a special health care need.

Because no other data source can provide this breadth of information for CSHCN at the national level and for all states, it is critical that these areas be addressed with sound research using the full armamentarium of techniques.

This presentation will examine the role and impact of qualitative evaluation, recognizing the need to carefully balance continuity and maintenance of trend analysis from the 2001 survey while simultaneously revising the instrument and addressing deficits. Qualitative tools such as focus groups, Technical Expert Panel discussions, key informant interviews, observation of field pretest interviews, and interviewer and supervisor debriefings highlighted key strengths and weaknesses of the instrument and methodology. The various stages of qualitative assessment complemented, but did not replace, quantitative analyses. Although qualitative work may be considered by some to be subjective and not as rigorous as quantitative methods, useful information can be gleaned by employing these techniques. Qualitative tools were employed to elicit the perceptions of survey methodologists, researchers, and parents of CSHCN, and bolster our understanding of what needs to be changed in the 2005–2006 survey. Pretesting in late 2004 will allow us to determine whether modifications implemented as a result of qualitative findings produced an improved instrument and methodology and a stronger overall survey.

CONCURRENT SESSION VIII-B:

PRETESTING AND USE OF NON-ENGLISH LANGUAGE QUESTIONNAIRES

Chair: Florina Serbanescu (Centers for Disease Control and Prevention, USA)

Adapting Cognitive Interview Techniques for Use in Pretesting Spanish Language Survey Instruments

Patricia Goerman (U.S. Census Bureau)

Pretesting methods that are currently used in survey research to develop and improve survey questions have been evolving for the past 20 years. While there has been increased focus on issues related to multicultural and multilingual survey design in recent years, (e.g. Harkness, Van de Vijver, and Mohler (eds.) 2003) there has been relatively little research on the appropriateness of specific pretesting methods with respondents from different cultural and/or linguistic groups.

Previous research has shown that some respondents have a great deal of difficulty with common cognitive interview probes and techniques. English-speaking respondents with low educational levels have been shown to have difficulty with both paraphrasing and think aloud probes (Willis, 2005; Bickert & Felcher, 1996; Wellens, 1994). Both Spanish and Chinese-speaking respondents have been shown to experience even greater difficulty with translated cognitive interview probes and techniques, including the use of think aloud, paraphrasing, process oriented probes, and meaning oriented probes (Pan, 2004; Carrasco, 2003; Potaka and Cochrane, 2002, 2004; Coronado and Earle, 2002; Blumberg & Goerman, 2000; Kissam, et al., 1993). In some cases these difficulties have been observed regardless of the educational level of respondents.

This paper presents an approach for continuing the process of tailoring and refining cognitive interview techniques to enhance their effectiveness across cultural and linguistic groups, with a focus on Spanish-speaking respondents. The paper discusses a plan of research for:

- Evaluating common cognitive interview techniques and probes translated into Spanish.
- Developing and testing alternative cognitive interview techniques for use with Spanish-speaking respondents.

Finally, the paper presents preliminary results and findings based on an implementation of this approach.

Development of Guidelines on the Use of Interpreters in Survey Interviews
Yuling Pan (U.S. Census Bureau)

Not much attention has been paid to the ways in which interpreters are currently being used in the conduct of federal household surveys. Many of these surveys entail personal interviews and are conducted largely by monolingual, English-speaking field interviewers. When a monolingual English-speaking field interviewer encounters a household in which the adult members speak little or no English he or she must rely on someone (an interpreter) who speaks the target language of the respondent for assistance in conducting the interview. Given the increase in the size of the non-English speaking population in the United States, ways in which interpreters are used, their training, and their qualifications for playing this role should be of key interest to federal statistical agencies that take pride in the quality of the survey data they collect.

This paper discusses the role of interpreters in federal survey interviews, and shows the need to develop guidelines for selecting, training, and evaluating interpreters to ensure the quality of data collected from non-English speaking households. This study is based on results from a Web search of best practices in using interpreters by survey organizations around the world and on observations of Census Bureau field staff conducting interviews with non-English speaking respondents. The aim of this study is:

1) To identify issues and challenges involved in conducting interviews (in English using interpreters and in languages other than English) with non-English speaking respondents;
2) To suggest ways to address the observed challenges associated with the conduct of in-person interviews using interpreters;
3) To assess what training is needed for interpreters to effectively interpret interviews with non-English speaking households.

The findings from this research will ultimately be used to develop Census Bureau guidelines for the use of interpreters in the conduct of household surveys.

The Use of Cognitive Interviewing and Behavior Coding to Evaluate Non-English Language Survey Questions: Lessons Learned
Gordon Willis (National Cancer Institute, USA)

Methods for the development and pretesting of survey questions in languages other than English are undergoing significant evolution. In particular, researchers are increasingly interested in adapting pretesting methods such as cognitive interviewing and behavior coding. It is not a simple matter, however, to apply these methods in the cross-cultural domain. In particular, extension to multiple languages poses particular challenges in terms of staffing, interpretation of results, and data analysis. We review three projects that applied cognitive interviewing to assess cross-cultural equivalence of health questions across English, Spanish, and Asian languages. Based on this variety of experiences we present selected results indicating how cognitive interviewing can be used to make conclusions about the functioning of survey instruments across languages, and make recommendations concerning future practice in this area.

Old Friends and Old Problems: Implementing Frequently-Used Question Types in the Cross-National Context
Janet Harkness (ZUMA, Germany)

Survey researchers know some questions are "bad news"—either difficult for respondents or difficult for the data collectors and analysts. Questions on household composition, income and ethnicity are familiar examples in the U.S. context. Various strategies continue to be investigated to address collecting better and appropriate data on these topics. In the cross-national context these questions often raise comparable problems as well as comparability issue, multiplied by the number of cultures we are trying to accommodate.

However, in cross-national studies we are also often confronted with difficulties we did not expect—principally because in the source questionnaire context the questions are not a problem. The paper looks at a number

of expected and unexpected problem questions pointing to the issues they raise for preparing cross-national instruments, adapting existing questions and administering instruments across countries. Examples are taken from questions on household composition and alcohol consumption, but also from a mental alertness test used for different purposes including procedures to decide whether to continue an interview.

Issues of address provision, spelling, memorizing phrases, and provision of information on geographical and temporal locations are presented, and the impact of cultural conventions and frameworks on required task completion discussed. The paper aims to address both micro and macro development and testing issues for extensively used questions.

CONCURRENT SESSION IX-A:

USE OF INNOVATIVE TECHNOLOGIES

Chair: Shawna Waugh (Energy Information Administration, USA)

Web-Based Factorial Survey and Statistical Analysis Software

Sam Addala and Vin Addala (e4xchange Corporation, USA), and Matthew Hogben (Centers for Disease Control and Prevention, USA)

Virtually anyone seeking information from a given group of people faces this dilemma: as one asks for more information, potential participants become more reluctant to provide all the information. Such reluctance is as likely to be on the basis of time and trouble as on grounds of sensitivity. Factorial surveys, however, permit researchers to sample different combinations of information from different respondents, thus allowing different respondents to contribute data to a subset of the full survey. To date, existing factorial survey software does not allow a researcher naVve to programming to design a survey and sampling protocol, and gather data (via internet) under the same umbrella.

e4Xchange Corporation, a software development organization and a NASA Industry Partner, has developed an intuitive and user friendly software solution to address the end-to-end needs of researchers conducting factorial and non-factorial surveys for the Centers for Disease Control and Prevention. e4Xchange has completed the Phase I proof of concept and has moved on to Phase II, full scale development. Proof of concept entailed developing factorial algorithms such that the survey designer could specify dimensions and levels of dimensions, which would then be sorted into all possible combinations and randomly presented.

Phase II is being built around four functions: design, Web publishing, response, and analysis. Design is the cardinal function, with the user able to develop "static" survey questions with varied response options, as well as factorial questions with varying numbers of factors and levels of factors. The designer simply writes a story (vignette) in a text box and names factors directly within that vignette. The designer then defines and names levels of factors, earmarks invalid combinations for partial factorial designs, and lets the program do the rest. Eligibility criteria and block randomization, if needed, can be built into the design, and the designer can specify or randomize the order of questions (or even just some of them), as well as denote skip patterns. Notably, the software is unique in that the user need not be a computer programmer to build complex factorial surveys.

Respondents log on, anonymously or confidentially, and either as a convenience sample or according to user-designed sampling protocols. The output is readable in a spreadsheet format, and therefore via common statistical programs (e.g., SPSS or SAS). Potential statistical analysis of output ranges from descriptive statistics to experimental frameworks from the General Linear Model (e.g., analysis of variance/regression, curve esti mation). The program has Web publication functions: when to publish surveys on the Web, for how long, and for how many respondents. Survey access can be Web-based, e-mail, or hard copy.

In sum, the software is designed to accommodate the needs of data-gathering and statistical analysis requirements of government, educational, and private industry.

Protecting Sensitive Tabular Data By Complementary Cell Suppression—Myth & Reality

Ramesh Dandekar (U.S. Department of Energy)

Title 13, U.S.C., Section 9 and the newly adopted CIPSEA of 2002 impose heavy financial fines and prison penalties for a public disclosure of sensitive statistical information. Currently, complementary cell suppression procedures are widely used by statistical agencies to protect sensitive tabular data from disclosure. It is generally believed that the complementary cell suppression procedures offer the best protection from wrongful disclosure of statistical information. In recent years LP-based automated audit procedures have been advocated and are being used to ensure the adequacy of protection offered by cell suppression patterns. LP-based lower and upper bounds for suppressed tabular cells are typically used to determine the adequacy of disclosure control measures. This paper identifies limitations of conclusions drawn using LP-based audit procedures. We utilize

commonly used analytical procedures to demonstrate the relative ease with which statistical disclosure of sensitive tabular data could occur. We conclude by providing additional safe guard measures required to avoid such disclosures.

Digital Capture of Geographic Feature Data for Surveys

Sarah Nusser (Iowa State University, USA)

In some surveys, it is of interest to record geospatial data such as housing unit coordinates, road networks, agricultural field boundaries, and water bodies as part of the survey data collection process. These shape data (i.e., points, lines, polygons) are inherently different from coded and fixed format questionnaire responses. Geographic information system (GIS) software can be used to record shape data, but the differences between a GIS software environment and that of a standard computer-assisted survey instrument present numerous challenges. Questions arise in balancing the linear structure of protocol enforcement with the iterative interactions that occur when working with geographic information; how to represent and configure feature editing environments for unskilled users; and whether constraints should be placed on the digitizing process to balance the accuracy of boundaries with the time spent delineating the feature. GIS software environments for survey data collection were explored via investigations with two agencies in the US Department of Agriculture. One study was conducted with National Agricultural Statistics Service field interviewers, who had not previously been exposed to computer-assisted data collection. Enumerators used pen-based tablet computers and a simplified ESRI ArcView application to delineate field boundaries on digital photographs during interviews with farm operators. Data were collected from observers and enumerators on the interview interaction and software use. In a second study, Natural Resource Conservation Service staff were trained in the use of GIS, but varied in the depth of their experience with GIS software. Participants delineated water bodies on high-resolution photographs using a custom ESRI ArcView application on desktop computers. The goals were to investigate the properties of area measurements at different scales in relation to work effort and to gain information on how users approached the delineation task in this environment through observation and digital capture of software sessions. We present results from these studies and on-going software development efforts.

Guaranteed Controlled Rounding for Many Totals in Multi-Way and Hierarchical Tables

Gordon Sande (Sande & Associates, Canada)

The problem of publishing a column of percentages so that its sum is 100 percent is an old problem. Sometimes a column of rounded entries will sum to the rounded total but often it will not. The problem is present for the common deterministic rounding rules of rounding down, rounding to nearest or rounding up. Various practical schemes for solving the problem of controlling the rounding of entries to match the rounded total by using a mix of rounding up and rounding down can be constructed. The rounding mode would no longer be the same for all entries. These schemes can be extended to include columns with internal subtotals, also called hierarchical structures. The problem has also been solved for simple two-way tables where the rounding of the entries must be chosen so that the rows and columns sum to their rounded totals.

The operations research theory for the two-way problem shows that the solution also applies when one of the dimensions has hierarchical structure. The results do not extend to cover the cases of two dimensions both with hierarchical structure or of more than two dimensions. The results apply when the subtotals can be arranged to form a network. To find a network for the two way table we would form a summation tree for the total where we first sum by rows to form row subtotals and then by columns to form the total. A summation tree is an abstract description of the mechanical tabulation method of break totaling used when subtotals are present. A second summation tree would be formed by summing first by columns and then by rows. A network can be formed by using one summation tree to go from the total to the internal entries and then the other summation tree to go from those entries to the total. A summation tree can be given for any number of dimensions with possible hierarchical structure in each dimension. Many different summation trees can be given and various of them will be required to determine the various subtotals. Not all summation trees will be required as the same subtotal may be present in different summation trees. An orderly description of all of the summation trees and subtotals is provided by the use of a lattice of internal entries and subtotals.

To address the general problem we would use the lattice to identify two summation trees with no common subtotals to provide a network. The network permits a controlled rounding of the internal entries, many subtotals and the total. The subtotals not in the network will not be part of the supplied solution so their values as reconstructed from the internal entries may not be the intended rounded values. The amount of adjustment away from the intended values of the non-network subtotals can be treated as a piecewise linear objective function for the possible rounding solutions which the network provides. Operations research techniques can be applied to determine optimal solutions for this objective function. In many cases the result will be that no adjustment is required as all subtotals can be successfully rounded. In the other cases the choice of the summation trees provides some user control over which subtotals will be guaranteed to have a controlled rounding.

When a complete successful rounding is not possible we see that some adjustments will be up and others will be down. The adjustments must balance as the total is rounded without adjustment as it is part of any network. Much as we occasionally choose to round up we may choose to only adjust up. This is a change in the problem statement which would be justified if we find that the solutions of the conventionally stated problem are unacceptable. The alternate formulation provides less control over which subtotals might be adjusted even though all are adjusted in the same direction. For example, the total might be adjusted up where otherwise it was guaranteed to be correctly rounded. The resulting problem can be addressed by operations research methods although the specialized techniques permitted by the presence of networks are no longer available.

The current constructive method of finding a random controlled rounding for a simple two-way table is a variation on the pivotal exchange of the primal simplex algorithm and can be readily generalized.

The more extensive analysis of the structure of the multi-way and hierarchical tables allows us to use more operations research theory in solving this problem.

CONCURRENT SESSION IX-B:

IMPROVING RESPONSE AND DATA QUALITY IN ESTABLISHMENT SURVEYS

Chair: Dennis Fixler (Bureau of Economic Analysis, USA)

Applying Knowledge of Business Surveys Response Processes to the Design of Data Collection Software at the U.S. Census Bureau

Amy Anderson and Rebecca Morrison (U.S. Census Bureau)

The U.S. Census Bureau gathers establishment data electronically for several surveys using data collection software. This system, called Surveyor, was initially created for the 2002 Economic Census, which was the first time an electronic reporting option was offered to all eligible business respondents. Surveyor was later adapted to collect information from other surveys during non-census years. In order to enhance the existing Surveyor system for the 2007 Economic Census, software developers requested detailed requirements. The Census Bureau augmented its own "lessons learned" from the 2002 Economic Census experience by developing investigative partnerships with business respondents in order to identify software requirements that meet the needs and expectations of these respondent users.

The goal of this research was to delve deeper into the business survey response process to obtain detailed descriptions of respondents' data collection tasks, such as the types of people involved and the location of data. After identifying these steps, we translated them into requirements for the software. Our goal is that, by working closely with actual respondents, we will enhance the software to dovetail with the respondents' established data collection routines.

This paper will describe how we gathered information about the data collection tasks business respondents engage in and how we applied that knowledge to develop requirements for software enhancement. We will also discuss the challenges we faced in creating requirements that would work for the many different types of response processes used by businesses.

The Response Process Model in Business Surveys: Lessons Learned by Using a Multi-Method Approach

Deirdre Giesen (Statistics Netherlands) and Tony Hak (Erasmus University, the Netherlands)

This paper describes the evaluation and redesign of the questionnaires for the Structural Business Survey Programme at Statistics Netherlands. This revision process encompasses 5 main phases. First, in the **problem finding** phase, the quality of the current questionnaires, in terms of response burden and data quality, was assessed by means of multiple evaluation methods. The results of this phase served as an input for the **diagnostic** phase, in which the results of the problem finding phase were validated with qualitative field research and likely causes of response burden and data error determined. These two steps resulted in recommendations for the revision of the questionnaires that were implemented in the **design** phase. Next, in the **test** phase, two versions of a new questionnaire were designed and tested. Finally, in the **decision** phase (Fall 2005), final decisions be taken on the new questionnaires which will be fielded in spring 2006. This paper reports on the first 4 phases of the SBS revision process and will conclude with our recommendations regarding content, design and structure of the SBS2005 questionnaire and the lessons learned with respect to the methods used in this project.

Improving Industry Descriptions for the Annual Refiling Survey

Karen Goldenberg and Monica Dashen (Bureau of Labor Statistics, USA)

Industry is a fundamental variable in government economic statistics, and the accurate classification of industry is important to ensuring high quality industry-based data. U.S. statistical agencies base industry on the North American Industry Classification System (NAICS). The sampling frame at the Bureau of Labor Statistics (BLS) includes a NAICS designation for each of the approximately 8.4 million business establishments, which it updates through the Annual Refiling Survey (ARS). In the ARS, respondents review a short description of the industry that corresponds to the assigned NAICS code, and indicate whether the description accurately reflects their establishments' main business activity.

The ARS industry descriptions consist of a broad descriptive statement, followed by an exhaustive or illustrative list of included and excluded items. Anecdotal reports from state agencies indicate that respondents are sometimes confused by these statements, especially the "exclude" statements, and report incorrectly as a result. In particular, state agency staffs tell BLS that respondents frequently reject descriptions because their business involves some level of an excluded item, even though the "includes" list covers their main activity. Respondents also reject correctly coded industry descriptions because they engage in an activity that is absent from an illustrative list. These employer misunderstandings result in additional state time being required to assign the correct industry code.

Our goal is to explore these State Agency reports in a controlled laboratory setting. We will provide respondents with a series of vignettes describing fictitious businesses and ask them to use the vignettes to complete the industry verification portion of ARS forms for those businesses. The vignettes will be based on common industries with large numbers of U.S. establishments. Some of the industry descriptions will be the versions currently used in the ARS, which contain both included and excluded items or activities. Others will be test versions that show only included items or activities. We will randomly assign current or test versions of each description so that respondents will see and work with both conditions, but for different industries. The vignettes will reflect real world situations and so may be ambiguous. Respondents will need to consult the descriptions in order to determine whether or not they accurately reflect the business described in the vignette. After respondents complete the ARS forms, we will debrief them on the reasoning they used to decide whether or not the description should be verified or rejected.

We will measure the frequency with which respondents correctly verify or reject the industry description based on the information in the vignette. We expect respondents to make more correct verifications for the test industry descriptions ("includes" only) than for the currently-used descriptions (with "include" and "exclude" statements). We expect respondents to incorrectly reject the currently-used industry descriptions more than the test descriptions because of the "exclude" statements. We will use the information from the debriefing interviews to understand how respondents reached their conclusion for each vignette.

Towards Reducing Error in an Establishment Survey Through Instrument Design: Identifying the Desired Navigational Path

Cleo Redline (National Science Foundation, USA)

A paper or Web self-administered survey instrument is really a physical object with many parts that need to work in unison for the express purpose of collecting information. Making a survey instrument work, at both the micro and macro level, such that the parts are transparent to the respondents and the respondents can accurately and efficiently operate the instrument (defined as performing the tasks in the order respondents are instructed to perform them) is a monumental undertaking. This is especially true in the case of an establishment survey with a large number of complicated parts, like the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS).

Expecting respondents to perform the tasks in the order they are instructed to perform them is another way of saying, "define the intended navigational path." By definition, the navigational path is under the control of the interviewer in an interviewer-administered survey, whereas this path is under the control of the respondent in a self-administered survey. Three important tasks that respondents must handle in a self-administered survey that would be under the control of the interviewer in an interviewer-administered survey are: 1) starting correctly, (2) reading and following instructions, definitions, or directions and (3) answering in terms of the correct reporting unit. The addition of these tasks in a self-administered instrument translates into greater complexity for respondents, which in turn translates into greater respondent burden and potential for error.

In the summer and fall of 2002, 12 in-depth cognitive/usability interviews were conducted in the DC Metropolitan area and 4 in Florida to learn more about how respondents navigated through and understood the GSS. The cognitive/usability interviews revealed that the GSS was really composed of three separate parts or instruments. Part 1 is meant to elicit a list of departments in science, engineering, or selected health fields from survey coordinators. Part 2 is forwarded to a respondent to collect the data at the departmental level. Part 3 is basically a database management tool meant to assist survey coordinators monitor (either their own or the departments') data collection process. Furthermore, 2 of these parts were administered both by paper and by Web. Thus, it became evident that there were 5 parts or instruments in need of improvement, not just one. It also became evident that the overarching problem with the survey was that respondents were being asked to carry out many tasks and that the overall design of the survey did not highlight and reduce competition for respondent's attention to one manageable task at a time—in other words, it did not help respondents navigate through the information correctly, beginning with getting started correctly.

This paper describes the many changes that were made to the GSS to induce respondents to start it correctly and to better understand the tasks they were being asked to perform in an effort to reduce what was discovered to be coverage error related to Part 1 and measurement error across all parts of the survey's instrument. It also attempts to extract the principles that were used to accomplish these objectives.

FCSM RESEARCH CONFERENCE AUTHOR AND COAUTHOR INDEX

NAME	SESSION
Natalie Abi-Habib	VI-B
Sam Addala	IX-A, Technical Demo
Vin Addala	IX-A, Technical Demo
Jeremy Aldworth	VI-B
Ingra Allred	III-A, Technical Demo
Amy Anderson	IX-B
Don Anderson	IV-A
Margo Anderson	V-B
Nelson Andrews	IV-A
Tihomir Asparouhov	II-A, Technical Demo
Kimberly Aspinwall	VIII-A
Justin Bailey	Technical Demo
Kent Bailey	Technical Demo
Andrew Baldwin	IV-A
Marco Ballin	II-C
Robert Baskin	II-C
Carol Bateman	III-A
Nancy Bates	VI-B
Michael Battaglie	II-B
Rodney Baxter	VI-B
Emmanuel Benhin	VII-C
HélPne Bérard	IV-B
Olivia Blum	V-C
Katherine Bowman	II-B
John Boyle	III-A
Howard Bradsher-Fredrick	IV-B, VII-B
Brian Burke	IV-A
Catherine Burt	VII-B
John Bushery	VII-C
Adam Carle	II-B
David Cella	VII-A
William Chan	II-B
Ted Chang	IV-B
Baoline Chen	VI-A
Shijie Chen	VII-B
Moon Jung Cho	III-B
Eun-Ha Choi	IV-A
Jai Won Choi	II-C
Daniel Chow	II-A
Susan Ciochetto	III-A
Richard Clayton	V-C, VI-A
Steve Cohen	V-B
Heather Contrino	III-A
André Cyr	III-C
James Dahlhamer	VI-B
Ramesh Dandekar	IX-A
Monica Dashen	IX-B
Leandro D'Aurizio	VI-C
Alexander Davies	III-C
Elizabeth Dean	VIII-A
Jill Dever	VII-B
Alan Dorfman	III-B
Jean Dumais	V-A
John Eltinge	III-B
Lawrence Ernst	III-B

NAME	SESSION
Brian Evans	IV-A
Ivan Faiella	VI-C
Robert Fay	III-C
Ron Fecso	VII-B
Lynn Fisher	VI-A
Ralph Folsom	III-C
Martin Frankel	II-B
George Freeland	III-A
Nanthini Ganapathi	III-A
Jose Gaudet	VII-C
Daniel Geller	VIII-A
Pamela Giambo	II-B
Deidre Giesen	IX-B
Claude Girard	III-C
G. Alastair Glendenning	VII-A
Patricia Goerman	VIII-B
Karen Goldenberg	IX-B
Rebecca Granger	II-B
Patricia Gwartney	VIII-A
Elizabeth Hahn	VII-A
Janet Harkness	VIII-B
Tony Hak	IX-B
Phuc Ho	II-B
Matthew Hogben	IX-A, Technical Demo
Susan Hunter	II-B
Vincnet Iannacchione	VII-B
Stefano Iezzi	VI-C
Claude Julien	VII-A
Sang Kang	II-B
Moshe Kelner	III-B
Joel Kennet	II-B
Elizabeth Kim	IV-A
Sheryl Konigsberg	V-C
Phillip Kott	IV-B
John Kovar	VII-A
Partha Lahiri	III-B
Paul Levy	IV-A
Timothy Li	V-B
Michael Link	II-B
Lenka Mach	V-A
Tzahi Makovky	III-B
Lawrence Malakhoff	Technical Demo
Isabelle Marchand	VII-C
Tamara Martin	VIII-A
Paul Massell	II-B
Jan Matuszko	VII-C
Tucker McElroy	II-A
Pamela McGovern	VII-C
Kristin Miller	Technical Demo
Thomas Moerhle	III-B
Ali Mokdad	II-B
Rebecca Morrison	IX-B
Charlotte Mueller	VI-A
Lynda Mulvany	III-A
Elizabeth Murphy	III-A

NAME	SESSION
Joe Murphy	VI-B
Bengt Muthen	II-A
Susan Myers	III-A
Balgobin Nandram	II-C
Andrea Neri	VI-C
Ronit Nirel	III-B
Sarah Nusser	IX-A
Benita O'Colmain	III-B
Kathleen O'Connor	VIII-A
Amy O'Hara	VI-C
Jim O'Reilly	Technical Demo
Laura Ozcoskun	V-A
Steven Paben	III-B
Dicy Painter	II-B
Yuling Pan	VIII-B
Swamy Paravastu	III-C
John Perry	VII-C
Andrea Piani	VI-B
Chester Ponikowski	III-B
Mark Pont	VII-C
Stuart Pursey	IV-B
Eric Rancourt	IV-B
Cleo Redline	IX-B
Debra Reed-Gillette	Technical Demo
Aryeh Reiter	III-B
Lauriane Robidou	V-A
Edwin Robison	III-C
Jean-François Rodrigue	VII-A
Pamela Rogers	V-C
Pedro Saavedra	III-B, VII-A
Akbar Sadeghi	VI-A
Gordon Sande	IX-A, Technical Demo
Mauro Scanu	II-C
Michael Schwerin	VIII-A
Lydia Scoon-Rogers	VI-C
William Seltzer	V-B

NAME	SESSION
Yann-Yann Shieh	V-C
Catherine Simile	VI-B
Avinash Singh	III-C, VII-B
Santokh Singh	IV-A
Lucien Smith	II-B
Marla Smith	VII-C
Paul Smith	VII-C
Miron Straf	Keynote Speaker
Barbara Stussman	VI-B
Michael Sverchkov	III-B
David Talan	V-C, VI-A
Lucilla Tan	II-B
Pam Tate	VII-C
Adrian Thibodeau	II-A
Gavin Thompson	VII-A
Katherine Thompson	V-A
Julie Trépanier	VII-A
Timothy Triplett	VI-B
Claude Turmelle	VII-A
Alfred Tuttle	V-B
Akhil Vaish	III-C
Paola Vicard	II-C
Andrey Vinokurov	VIII-A
Kevin Wang	VI-B
William Weinig	VII-B
Paula Weir	III-B
Quatracia Williams	V-A
Diane Willimack	V-B
Gordon Willis	VIII-B
Hoke Wilson	VII-A
Jeff Wilson	II-A
John Wohlford	VI-A
David Woodwell	VII-B
Yong You	III-C
Peter Zadrozny	VI-A
Tamara Zimmerman	III-C

Reports Available in the Statistical Policy Working Paper Series

1. Report on Statistics for Allocation of Funds, 1978 (NTIS PB86-211521/AS)
2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques, 1978 (NTIS PB86-211539/AS)
3. An Error Profile: Employment as Measured by the Current Population Survey, 1978 (NTIS PB86-214269/AS)
4. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics, 1978 (NTIS PB86-211547/AS)
5. Report on Exact and Statistical Matching Techniques, 1980 (NTIS PB86-215829/AS)
6. Report on Statistical Uses of Administrative Records, 1980 (NTIS PB86-214285/AS)
7. An Interagency Review of Time-Series Revision Policies, 1982 (NTIS PB86-232451/AS)
8. Statistical Interagency Agreements, 1982 (NTIS PB86-230570/AS)
9. Contracting for Surveys, 1983 (NTIS PB83-233148)
10. Approaches to Developing Questionnaires, 1983 (NTIS PB84-105055)
11. A Review of Industry Coding Systems, 1984 (NTIS PB84-135276)
12. The Role of Telephone Data Collection in Federal Statistics, 1984 (NTIS PB85-105971)
13. Federal Longitudinal Surveys, 1986 (NTIS PB86-139730)
14. Workshop on Statistical Uses of Microcomputers in Federal Agencies, 1987 (NTIS PB87-166393)
15. Quality in Establishment Surveys, 1988 (NTIS PB88-232921)
16. A Comparative Study of Reporting Units in Selected Employer Data Systems, 1990 (NTIS PB90-205238)
17. Survey Coverage, 1990 (NTIS PB90-205246)
18. Data Editing in Federal Statistical Agencies, 1990 (NTIS PB90-205253)
19. Computer Assisted Survey Information Collection, 1990 (NTIS PB90-205261)
20. Seminar on Quality of Federal Data, 1991 (NTIS PB91-142414)
21. Indirect Estimators in Federal Programs, 1993 (NTIS PB93-209294)
22. Report on Statistical Disclosure Limitation Methodology, 1994 (NTIS PB94-165305)
23. Seminar on New Directions in Statistical Methodology, 1995 (NTIS PB95-182978)
24. Electronic Dissemination of Statistical Data, 1995 (NTIS PB96-121629)
25. Data Editing Workshop and Exposition, 1996 (NTIS PB97-104624)
26. Seminar on Statistical Methodology in the Public Service, 1997 (NTIS PB97-162580)
27. Training for the Future: Addressing Tomorrow's Survey Tasks, 1998 (NTIS PB99-102576)
28. Seminar on Interagency Coordination and Cooperation, 1999 (NTIS PB99-132029)
29. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 1999 (NTIS PB99-166795)
30. 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings, 2000 (NTIS PB2000-105886)
31. Measuring and Reporting Sources of Error in Surveys, 2001 (NTIS PB2001-104329)
32. Seminar on Integrating Federal Statistical Information and Processes, 2001 (NTIS PB2001-104626)
33. Seminar on the Funding Opportunity in Survey Research, 2001 (NTIS PB2001-108851)
34. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2001 (NTIS PB2002-100103)
35. Seminar on Challenges to the Federal Statistical System in Fostering Access to Statistics, 2004
36. Seminar on the Funding Opportunity in Survey Research, 2003
37. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2003
38. Summary Report of the FCSM-GSS Workshop on Web-Based Data Collection, 2004

39. Seminar on the Funding Opportunity in Survey Research, 2004 (Forthcoming)
40. Seminar on Achieving Statistical Quality in Diverse and Changing Environments (Forthcoming)
41. Seminar on the Funding Opportunity in Survey Research, 2005 (Forthcoming)
42. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2005 (Forthcoming)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161; telephone: 1-800-553-6847. The Statistical Policy Working Paper series is also available electronically from FCSM's Web site <http://www.fcsm.gov>.

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

Issued November 2005